

N° d'Ordre : 1 (1975)

N° CNRS : AO 10 826

THESE

présentée devant l'UNIVERSITE de NANTES

pour obtenir le grade de

DOCTEUR ès-SCIENCES

par

Alain BARRAUD

ECOLE NATIONALE SUPERIEURE DE MECANIQUE



REALISATION MINIMALE ET APPROXIMATION OPTIMALE
DES SYSTEMES DYNAMIQUES LINEAIRES INVARIANTS

Soutenue le 17 janvier 1975 devant la Commission d'Examen

M. MEZENCEV Président

M. de LARMINAT

M. FOULARD

M. BRILLOUET

M. RAYMOND

Examineurs

THESE

présentée devant l'UNIVERSITE de MANTES

pour obtenir le grade de

DOCTEUR ès-SCIENCES

par

Alain BARRAUD

ECOLE NATIONALE SUPERIEURE DE MECANIQUE



REALISATION MINIMALE ET APPROXIMATION OPTIMALE
DES SYSTEMES DYNAMIQUES LINEAIRES INVARIANTS

Soutenue le 17 janvier 1975 devant la Commission d'Examen

M. MEZENCEV Président

M. de LARMINAT

M. FOULARD

M. BRILLOUET

M. RAYMOND

} Examinateurs



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41555012_0014

UNIVERSITE DE NANTES

Président : M. KERNEIS - Médecine -

Vice-Président : M. SPARFEL - Pharmacie -

Liste des enseignants des disciplines scientifiques :

Professeurs	M. BRILLOUET	Mathématiques
	M. PAUC	«
	M. PETRESCO	«
	M. RAFFIN	«
	M. PIRONNEAU (ENSM)	Mécanique
	M. RAVILLY (ENSM)	«
	M. TARDY (ENSM)	«
	M. AVIGNON	Physique
	M. BOULLLOUD	«
	M. COLLET	«
	M. MINN	«
	M. MEZENECV (ENSM)	«
	M. MOUTON	«
	Melle OFFRET (ENSM)	«
	M. DUNOYER	Chimie
	M. GRAFF	«
	M. MARTIN	«
	M. PIERI	«
	M. QUINIOU	«
	M. ROUXEL	«
	M. WOJTKOWIAK	«
	M. GARRIGUES	Biologie
	Mme LE DOUARIN	«
	M. LE DOUARIN	«
	M. SELLIER	«
	M. CAVET	Sciences de la Terre
	M. FORESTIER	«
Professeurs associés	M. BRARD (ENSM)	Mécanique
Professeurs sans chaire	M. FOUQUES	Mathématiques
	M. IBISCH	«
	M. LEBORGNE	«

Professeurs sans chaire	M. BARDON	Physique
	M. CAILLER (IUT Nantes)	«
	M. GOUREAUX	«
	M. LASKAR	«
	M. LE NAOUR	«
	M. ROSIER	«
	M. BOUCHY (IUT Nantes)	Chimie
	M. CHABANEL	«
	M. GOBIN (ENSM)	«
	M. POINTEAU (ENSM)	«
	M. TOURNOUX	«
	M. DUPONT	Biologie
	M. SAUDRAY	«
	M. OTTMANN	Sciences de la Terre
Maîtres de Conférences	M. DHOMBRES	Mathématiques
	M. DO (ENSM)	«
	M. GAPAILLARD	«
	M. GATESOUBE	«
	M. LAMEAU (ENSM)	«
	M. PHAN-THE-LAI (Chargé d'Enseignement)	«
	M. VOGEL	«
	M. DUBIGEON (Chargé d'Enseignement IUT)	Mécanique
	M. GUEVEL (ENSM)	«
	M. SULMONT (ENSM)	«
	M. VAUSSY (ENSM)	«
	M. VILAINE (ENSM)	«
	M. LE BIHAN	Physique
	M. LE CORRE (ENSM)	«
	M. LE MAITRE	«
	M. LONCKE	«
	M. THOMAS (IUT Nantes)	«
	M. LE MAITRE (ENSM)	Chimie
	Mme MARTIN	«
	M. AMOURIQ	Biologie
	M. LEBRUN	«
	Mme NICOLAS-PRATT	«
	M. NICOLAS	Sciences de la Terre

Secrétaire Général : M. LERAY

*L'approximation, c'est l'objectivation inachevée...,
à la fois consciente de son insuffisance et de son progrès.*

G. BACHELARD

A mes fils,

Laurent

et

Thomas

AVANT PROPOS

La présente thèse a été entreprise au Laboratoire d' Automatique de l'ECOLE NATIONALE SUPERIEURE DE MECANIQUE de Nantes que dirige Monsieur le Professeur MEZENCEV.

Qu'il trouve ici l'expression de ma profonde reconnaissance pour l'intérêt qu'il n'a cessé de porter à cette étude et la confiance qu'il m'a accordée en me chargeant de l'implantation, au Laboratoire, d'une Bibliothèque de Programmes à la mesure des problèmes que nous étions tous amenés à traiter.

Je ne saurais dire toute ma gratitude à mon Ami et Collègue, Monsieur de LARMINAT, initiateur de ce travail, auprès de qui j'ai toujours trouvé des conseils inestimables. Qu'il sache que mon souhait le plus vif est que cette collaboration fructueuse se poursuive longtemps encore.

Je voudrais également exprimer toute la reconnaissance que je garde à mon parrain du C.N.R.S., Monsieur le Professeur FOULARD de l'Institut Polytechnique de Grenoble, dont les avis hautement autorisés ont été de précieux guides dans la réalisation de ce mémoire.

J'adresse mes sincères remerciements à Monsieur le Professeur RAYMOND du Conservatoire National des Arts et Métiers, d'avoir accepté de participer à ce Jury. Qu'il me soit permis ici, de lui dire combien je lui sais gré de l'intérêt qu'il m'a toujours témoigné.

Que Monsieur le Professeur BRILLOUET, Directeur du Centre de Calcul Universitaire, sache combien je suis sensible à l'honneur qu'il me fait en acceptant de juger ce travail. Nous devons tous, à son dynamisme, l'essor d'un Centre de Calcul, sans lequel je n'aurais personnellement pu mener à bien la partie expérimentale de mes recherches.

Je tiens également à exprimer mes remerciements à Monsieur CHEVRE pour toutes les facilités qu'il m'a octroyées lors de la mise en place de notre Bibliothèque de Programmes ("Package Harwell"). Je voudrais aussi remercier personnellement Madame d'HERVE et Monsieur SAULNIER pour le dévouement dont ils ont fait preuve, durant des journées entières, en faisant passer mes programmes.

Qu'il me soit donné ici l'occasion de remercier nombre de Collègues du Laboratoire d'Automatique avec qui j'ai eu de fructueux échanges, je pense particulièrement à Messieurs ELLOY et CALVEZ sur le plan informatique, Messieurs THOMAS, SARLAT et DESCUSSE en ce qui concerne les problèmes de commande, à Monsieur PIASCO pour sa connaissance de la Centrale de Cheviré, dont il m'a fait profiter, et aussi à Monsieur LAFAY pour sa contribution à certains tests numériques en programmation non linéaire.

Enfin, que Mesdames LEMAIRE et POTIRON, Mesdemoiselles CORBIN et LECOQ, trouvent ici tous mes remerciements pour le courage et la bonne volonté dont elles ont fait preuve en dactylographiant ce manuscrit. Je tiens également à remercier le Personnel de l'Imprimerie qui a participé à cette édition.

PRINCIPALES NOTATIONS

[5]	: report bibliographique en fin de Chapitre
(II-7)	: équation 7, Chapitre II
†	: renvoi en bas de page
∀	: quel que soit
∃	: il existe
$\triangleq$	: égal par définition
*	: opérateur convolution (continu)
★	: opérateur convolution (discret)
s	: variable de Laplace
$\mathcal{L}(\cdot)$	: transformation de Laplace
z	: variable de la transformée en Z
δ_t	: symbole de Kronecker
$\delta(t)$	: impulsion de Dirac
a, α	: scalaires
$\underline{v}$	: vecteur colonne
$\underline{v}^T$	: vecteur ligne (T transposition)
$ \cdot $	: norme (eudidienne sauf indication contraire)
A	: matrice
R(A)	: espace engendré par les colonnes de A
$\rho(A)$	: rang de A = dimension de R(A)
N(A)	: noyau de A
$\nu(A)$	: dimension de N(A)
A^{-1}	: inverse de A
A^+	: pseudo inverse de A
det A	: déterminant de A
1	: matrice unité
R^k	: espace vectoriel à k dimension sur le corps des réels
$U \xrightarrow{g} X$	: application de l'espace U vers l'espace X
$\left. \begin{array}{l} \underline{u}, \underline{u}(t), \underline{u}_t \\ \underline{u}(s), \underline{u}(z) \end{array} \right\}$	: vecteur d'entrée de dimension m
$\left. \begin{array}{l} \underline{y}, \underline{y}(t), \underline{y}_t \\ \underline{y}(s), \underline{y}(z) \end{array} \right\}$	: vecteur de sortie de dimension p

$G(s)$	: matrice de transfert en s
$G(t)$	: matrice des réponses impulsionnelles
$F(z)$	: matrice de transfert en z
F_t	: matrice des séquences de pondération
Σ_n	: système dynamique d'ordre n
$\Sigma(A,B,C,D)$	: système dynamique continu
$\Sigma(\Phi,\Gamma,H,D)$	: système dynamique discret
Σ^*	: système dynamique dual
$\Delta_t, \Delta_t(A,B), \Delta_t(\Phi,\Gamma)$	: matrice accessibilité ou de commandabilité
$\Lambda_t, \Lambda_t(A,B), \Lambda_t(\Phi,\Gamma)$	: matrice d'observabilité ou de restructibilité
$ \underline{e} $	: erreur d'équation
$ \underline{\varepsilon} $	: erreur de sortie

° °

°

SOMMAIRE

	<u>Pages</u>
<u>INTRODUCTION</u>	1
<u>Bibliographie</u>	7
 <u>CHAPITRE I</u> Définitions et propriétés des systèmes dynamiques linéaires invariants	 11
1 - <u>Introduction</u>	12
2 - <u>Description externe - relation d'entrée/sortie</u>	12
2-1. Systèmes continus	13
2-2. Systèmes discrets	14
3 - <u>Description interne - modèle d'état</u>	15
3-1. Systèmes continus	16
3-2. Systèmes discrets	17
4 - <u>Réalisation d'une relation d'entrée/sortie</u>	17
4-1. Réalisabilité	18
4-2. Classe d'équivalence	18
4-3. Généralisation du concept de réalisation	19
5 - <u>Conclusion</u>	20
<u>Bibliographie</u>	21
 <u>CHAPITRE II</u> Structure de l'espace d'état	 23
1 - <u>Introduction</u>	24
2 - <u>Préliminaires</u>	24
3 - <u>Commandabilité et accessibilité</u>	26
3-1. Critère d'accessibilité	27
3-2. Critère de commandabilité	30
3-3. Décomposition de l'espace d'état (1ère partie)	32

	<u>Pages</u>
4 - <u>Observabilité et restructibilité</u>	36
4-1. Critère d'observabilité	37
4-2. Critère de restructibilité	39
4-3. Décomposition de l'espace d'état (2ème partie)	41
5 - <u>Structure de l'espace d'état définie par l'accessibilité et l'observabilité</u>	43
6 - <u>Réalisation minimale d'un système dynamique</u>	45
6-1. L'approche classique	45
6-2. Un formalisme plus général	46
7 - <u>Conclusion</u>	49
<u>Bibliographie</u>	51
 <u>CHAPITRE III</u> Formes canoniques et structure invariante	 53
1 - <u>Introduction</u>	54
2 - <u>Préliminaires</u>	54
3 - <u>Systèmes monovariabiles</u>	60
4 - <u>Systèmes mono-sortie et mono-entrée</u>	68
4-1. Systèmes mono-sortie	68
4-2. Systèmes mono-entrée	70
5 - <u>Définition d'une structure pour l'opérateur (A,B)</u>	72
5-1. Définition d'un système complet d'invariants indépendants	72
5-2. Découplage de l'opérateur (A,B)	79
5-3. Généralisation au cas où $\rho[\Delta_n(A,B)] < n$	83
6 - <u>Formes canoniques multivariabiles</u>	84
6-1. Formes canoniques accessibles	85
6-2. Formes canoniques observables	88
7 - <u>Conclusion</u>	90
<u>Bibliographie</u>	91

	<u>Pages</u>
<u>CHAPITRE IV</u> Réalisations canoniques minimales : Algorithmes	93
1 - <u>Introduction</u>	94
2 - <u>Réalisation d'une matrice de transfert rationnelle propre</u>	95
2-1. Algorithme 1 (forme observable)	96
2-2. Algorithme 2 (forme accessible)	102
3 - <u>Réalisation d'une suite de paramètres de Markov</u>	104
3-1. Algorithme 3 (forme observable)	105
3-2. Algorithme 4 (forme accessible)	106
4 - <u>Réalisation directe d'entrées/sorties quelconques (cas discret)</u>	107
4-1. Relations d'entrée/sortie associées à la forme canonique observable (2)-(4)	108
4-2. Relations d'entrée/sortie associées à la forme canonique observable (1)-(3)	114
4-3. Identifiabilité et réalisabilité	117
4-4. Algorithmes de réalisation	129
5 - <u>Réalisation directe d'entrées/sorties quelconques (cas continu)</u>	132
5-1. Relation d'équivalence entre système continu et discret	134
5-2. Détermination du modèle continu dans la forme canonique observable	137
6 - <u>Conclusion</u>	143
<u>Bibliographie</u>	144
<u>CHAPITRE V</u> Approximations optimales - Algorithmes - Résultats expérimentaux	147
1 - <u>Introduction</u>	148
2 - <u>Approximation des systèmes monovariables</u>	149
2-1. Définition de critères d'approximation	152
2-2. Minimisation de l'erreur de sortie	156
2-3. Estimation de l'ordre du système	164

- 3 - Résultats expérimentaux (systèmes monovariabiles)
 - 3-1. Etude comparative de différentes techniques de minimisation de l'erreur de sortie
 - 3-2. Identification d'un système simulé
- 4 - Approximation des systèmes multivariabiles
 - 4-1. Une approche par erreur d'équation
 - 4-2. Minimisation de l'erreur de sortie
 - 4-3. L'algorithme mixte
- 5 - Résultats expérimentaux (systèmes multivariabiles)
 - 5-1. Description de la tranche 5
 - 5-2. Définition des grandeurs d'entrées/sorties
 - 5-3. Identification de la tranche 5
- 6 - Commentaire sur les méthodes numériques
- 7 - Conclusion

Bibliographie

CONCLUSION

A la différence de la théorie classique des asservissements, fondée sur la connaissance d'une représentation *externe* du système, les techniques modernes de commande et de filtrage s'appuient sur la description plus fine des processus que définit leur représentation *interne*, c'est-à-dire les modèles dits d'*état*. On conçoit donc l'importance attachée aux méthodes de recherche de tels modèles, problème défini comme étant celui de la *réalisation*.

Un bref rappel historique sur ces problèmes de réalisation nous permettra d'abord de situer, dans leur contexte global, les idées que nous avons voulues dégager, concernant un type de solution qui semble devoir s'imposer et dont nous donnerons quelques applications.

L'essor de ce domaine date, en fait, de la théorie générale des systèmes de commande développée par KALMAN en 1960 [1]. L'importance des résultants acquis dans le cadre de la réalisation minimale des systèmes linéaires, nous invite à y distinguer plusieurs branches suivant, par exemple la nature des *données* du problème : matrices de transfert, paramètres de MARKOV, opérateurs différentiels (aux différences), ou encore ensemble d'entrées-sorties synchrones quelconques.

Une des premières solutions (1963) a été proposée par GILBERT [2] concernant la réalisation de matrices de transfert n'ayant que des pôles simples. KALMAN [3] a généralisé ce résultat un peu plus tard (1965) en élaborant un algorithme basé sur la forme canonique de SMITH des λ matrices (pôles de multiplicité quelconque). Cette procédure, faisant appel à des manipulations de matrices polynomiales, donc peu adaptée au traitement numérique est difficile à mettre en oeuvre. Il en va de même pour la méthode que l'on trouve, plusieurs années après dans l'article d'HEYMANN [4] essentiellement théorique.

Dans une voie nettement différente, ROSENBROCK [5] a développé dans le contexte de sa représentation particulière des systèmes, un algorithme très simple à programmer, car basé sur des opérations matricielles élémentaires [6]. Celui-ci consiste, en fait, à extraire la partie accessible ou/et observable d'une réalisation triviale [7]. L'algorithme de MAYNE, relevant d'un même esprit, présente, quant à lui, l'intérêt majeur, de faire apparaître une structure du système [8], [9]. La forme obtenue ici (blocs compagnons couplés) est extrêmement intéressante, à notre sens, parce que relativement

simple à exploiter quel que soit le type de données, et en particulier beaucoup plus facile à construire que la forme canonique de JORDAN, à laquelle de nombreux auteurs se sont néanmoins attachés. Certains, comme CAINES [10] [11] se sont limités au cas où l'espace d'état est cyclique, arguant du fait que tout espace peut être décomposé en de tels sous-espaces. D'autres approches ont été suggérées, soit en passant par l'intermédiaire d'une réalisation non minimale [12], [13], [14], soit directement [15], [16]. Quoiqu'il en soit ces procédures n'en demeurent pas moins lourdes et supposent implicitement connus les pôles du système.

Parallèlement à ces développements, le problème a été abordé en prenant comme données la suite (théoriquement infinie) des paramètres de MARKOV du système. L'algorithme de HO [17] [18] est à l'origine de nombreuses techniques reposant sur les propriétés des matrices de HANKEL du système [19] [20] [21]. Nous distinguerons, de celle-ci, la solution donnée par ACKERMANN [22], d'une part simple à coder, et d'autre part, conduisant à une structure "blocs compagnons couplés". Il convient de noter, qu'indépendamment de toute méthodologie, on fait souvent abstraction de l'hypothèse concernant l'infinité de paramètres sensés être connus (et utilisés !). On résout donc la plupart du temps, ce que l'on appelle, sur le plan théorique, le problème de réalisation partielle [23] [24]. Il est cependant clair qu'une telle difficulté ne doit pas se présenter si ces paramètres sont issus de matrices de transfert rationnelles CHEN [25]. Cette remarque offre d'ailleurs une alternative pour réaliser ce type de modèle externe [26] résultat encore obtenu ici dans la forme "blocs compagnons couplés".

Un autre aspect du problème consiste à prendre directement comme données des couples d'entrées-sorties quelconque du système, sans passer par la phase d'identification classique que nécessite le plus souvent la connaissance des transmittances ou des paramètres de MARKOV. Dans ce domaine, les premiers résultats significatifs sont encore assez récents (GOPINATH) [27] (BUDIN) [28] [29]. Ces méthodes consistent à sélectionner, "*dans*" leur ordre naturel, parmi les lignes d'une matrice du type HANKEL, constituée d'entrées-sorties, celles qui sont linéairement indépendantes. Les matrices de la réalisation obtenue ne présentent alors aucune forme particulière, à l'exception de la matrice de sortie composée uniquement de 0 ou 1.

Les techniques que nous avons développées dans [30] [31] sont d'un esprit voisin puisque également basées sur la sélection de lignes ou colonnes linéairement indépendante extraite d'une certaine matrice d'entrée-sortie. La différence, dont les conséquences sont fondamentales, porte sur la manière dont est effectuée cette sélection. La façon dont on opère permet en effet de déterminer simultanément la structure d'une forme canonique (ici blocs compagnons couplés) et de déterminer les "invariants" du système, opération que GUIDORZI appelle "identification structurale" [34]. Rappelons qu'à la même époque, ACKERMANN [32] avait proposé indépendamment un autre algorithme conduisant au même type de structure. Cette approche commune est, à notre avis, plus efficace que celle développée peu après par GUIDORZI (et al.) [33] [34] dans la mesure où elle conduit à une paramétrisation plus simple de la forme canonique (blocs compagnons couplés). Notons d'une façon générale que ces algorithmes ne font, en aucun cas, appel à des réalisations intermédiaires non minimales qui rendrait le problème quelque peu trivial, puisque ramené à une suite d'identifications classiques monosortie (réalisations observables immédiates), suivie de l'élimination des états non accessibles [35].

Sans chercher à être exhaustif, signalons néanmoins pour terminer ce tour d'horizon quelques résultats obtenus dans le cadre de la réalisation des systèmes décrits par opérateurs différentiels (aux différences) POLAK [36], GUEGUEN [37] et CHEN [38]. En raison de l'emploi peu fréquent de ce type de modèle, nous ne reviendrons pas sur ce problème par la suite.

Pour résumer, et en complément de ce qui vient d'être dit, nous distinguerons donc essentiellement deux types d'approches : la réalisation d'un modèle externe (matrices de transfert, paramètres de MARKOV), et la réalisation d'un système caractérisé par des couples d'entrées-sorties. On notera d'autre part l'évolution significative vers la construction de réalisations minimales ayant une structure "blocs compagnons couplés".

Cette tendance définit en fait le thème central de notre contribution développée principalement au Chapitre IV. Avant de parvenir à ces résultats, il nous a paru nécessaire de situer avec précision le cadre de notre travail, ce qui nous a amenés à rappeler, au cours d'un chapitre introductif, un certain nombre de "Définitions et propriétés des systèmes dynamiques linéaires invariants". Un second chapitre a été consacré à l'étude de la "Structure de l'espace d'état" à l'occasion duquel, il nous a été permis de préciser dans le contexte des systèmes discrets, des résultats souvent épars, ou incomplets, concernant les concepts de commandabilité et d'accessibilité, ainsi que leurs duaux : restructibilité et observabilité. Vient ensuite un chapitre de synthèse, traitant d'un certain type de formes canoniques mono et multivariables, exposé sur la base de la notion de structure invariante et présenté comme le moyen de définir une paramétrisation intéressante des systèmes dynamiques. Cette approche constitue un élément fondamental à partir duquel seront développés, au chapitre IV les algorithmes proprement dits de réalisation minimales, puis au chapitre V ceux concernant les problèmes d'approximation.

Une part importante de notre contribution porte donc sur la démarche unifiée suivant laquelle seront élaborés ces algorithmes, conduisant, quel que soit le type de données, à une structure blocs-compagnons-couplés. Ces résultats acquis dans le cadre des systèmes discrets généralisent et rapprochent diverses solutions données indépendamment par plusieurs auteurs [8] [22] [26] [30] [32].

Compte-tenu du contexte dans lequel sont le plus souvent utilisés ces modèles (commande numérique), il ne paraît pas fondamental d'envisager des solutions analogues ayant trait à la détermination directe d'un modèle continu. Bien que ce soit possible, il nous paraît plus efficace dans le cas où un tel modèle se révélerait nécessaire de passer par l'intermédiaire d'un système discret dont il resterait ensuite à déduire un modèle continu "équivalent". C'est cette solution que nous avons développée à la fin du chapitre IV avant d'aborder le problème d'approximation, caractéristique de la modélisation des systèmes physiques, quels qu'ils soient, et dont l'importance est à la mesure des difficultés qu'il représente.

La différence qui existera par définition entre modèle et objet, est à l'origine d'une question préliminaire : selon quel critère l'un sera-t-il une "bonne" approximation de l'autre ? Il n'existe malheureusement pas de réponse générale à une telle question, tout au plus pourra-t-on préciser la qualité d'un modèle pour un objectif donné. Parmi les solutions que l'on peut envisager, on a donné une importance particulière à deux critères : l'"erreur d'équation" et l'"erreur de sortie" en raison de la simplicité que requière la mise en oeuvre du premier et la signification physique précise du second.

L'intérêt de combiner ces deux approches est évident. Ceci nous a conduit à traiter en détail la minimisation de l'erreur de sortie dont la solution ressort de la programmation non linéaire. Dans ce domaine, nous présentons des techniques récentes, qui, pensons-nous sont appliquées pour la première fois dans le contexte de l'identification. Les tests numériques que nous rapportons ici, et effectués en simulation sur des systèmes monovariables, montrent clairement ce qu'apportent de telles solutions.

Le cas des systèmes multivariables est nettement plus complexe à appréhender notamment en raison des difficultés numériques qu'il présente, mais aussi parce qu'il pose dans des termes nouveaux la définition d'une structure canonique.

Nous avons évoqué, à ce sujet, une solution que l'on pourrait prendre comme référence, mais dont la mise en oeuvre ne serait raisonnablement possible que pour des systèmes de petites dimensions. La solution effective à des problèmes d'application industrielle passe nécessairement par un compromis dont l'"algorithme mixte" (ch. V) est un exemple, qui s'est révélé à l'expérience extrêmement intéressant.

Nous voudrions signaler ici que le succès même de cette approche dépend en majeure partie du large développement que l'on a donné à l'analyse des problèmes d'un point de vue numérique. Précisons simplement que si ceux-ci étaient relativement "classiques" dans leur principe, ils n'en demeureraient pas moins difficiles à traiter, par leur taille : systèmes linéaires de plus de 1 000 équations et 100 inconnues, programmation non linéaire avec plusieurs dizaines de variables, par leur mauvais conditionnement dû à l'extrême sensibilité du critère d'"erreur de sortie" par rapport à certaines variables (pôles du modèle par exemple), etc... difficultés auxquelles il fallait ajouter un système informatique modeste (IBM 360-30, 64 K en 1972 ; 370-135, 96 k en 1973 et enfin 144 k + mémoire virtuelle de 400 K en 1974).

De ce fait le lecteur ne devra pas s'attendre à trouver ici une illustration complète de tous les points dont nous avons parlé en raison de l'effort d'analyse et de programmation que cela aurait représenté. En effet l'écriture, la mise au point et une exploitation même succincte des programmes relatifs à l'ensemble des algorithmes, que nous décrivons chapitre IV et V, auraient nécessité au moins le double du temps que nous avons déjà consacré pour réaliser ce mémoire.

Nous avons donc retenu, dans le cadre des résultats expérimentaux que les aspects essentiels de notre travail. Parmi ceux-ci, nous avons donné une place toute particulière au problème d'approximation multivariable à partir d'entrées/sorties quelconques. Sa généralité et les résultats obtenus à l'issue de campagnes d'essais réalisés sur une centrale thermique témoignent à la fois de l'intérêt des solutions que nous proposons et de leur efficacité face à des problèmes de dimensions réalistes.

-7-

BIBLIOGRAPHIE

A l'exception de la première référence d'intérêt général, les suivantes ont trait au problème de réalisation et sont classées par sujet et ordre de citation dans le texte, comme suit :

- [2] à [16]: réalisation d'une matrice de transfert
- [17] à [26]: réalisation de réponses impulsionnelles
- [27] à [35]: réalisation d'entrées sorties quelconques
- [36] à [38]: réalisation de systèmes différentiels.

- [1] R.E. KALMAN "On the general theory of control systems"
Proc. of the first IFAC Congress Moscow (1960) p 481-492
- [2] E.G. GILBERT "Controllability and observability in multivariable control systems"
SIAM J. Control Vol 2 n° 1, p 128-151 (1963)
- [3] R.E. KALMAN "Irreducible realisations and the degree of a rational matrix"
J. Soc. Indust. Appl. Math. Vol 13 n° 2 June 1965
p 520-544.
- [4] M. HEYMANN "Transfer equivalence of linear dynamical systems"
J.A. THORPE SIAM J. Control Vol 8-n° 1 Feb. 1970, p 19-40.
- [5] H.H. ROSENBROCK "Computation of minimal representation of a rational transfer function matrix"
Proc. I.E.E. Vol 115 n° 2, Feb 1968, p 325-327.
- [6] N. MUNRO "Minimal realization of transfer function matrices
R.S. MC LEOD using the system matrix"
Proc I.E.E., Vol 118 n° 9, sept. 1971, p 1298-1301
- [7] D.Q. MAYNE "An elementary derivation of Rosenbrock's minimal realization algorithm"
I.E.E.E. Trans. Auto. Cont. June 1973, p 306-307
- [8] D.Q. MAYNE "Computational procedure for the minimal realization of transfer function matrices"
Proc I.E.E. Vol 115 n° 9 sept 1968, p 1363-1368.
- [9] D.M. ROBERTS "A computer program for the minimal realization of transfer function matrices"
I.E.E.E. Trans. Aut. Cont. Aug. 1969, p 412-413.
- [10] P.E. CAINES "Minimal realization of transfer function matrices"
Int. J. Control Vol 13 n° 3 (1971) p 529-547.
- [11] P.E. CAINES "Unique state variable models with applications to the identification of linear systems"
Control System Centre Report n° 140 U,M,I,S,T.

- [12] C. GUEGUEN
E. TOUMIRE "Mise sous forme de variable d'état par la méthode matricielle"
Automatisme 13 (1968) p 199-203.
- [13] S.P. PANDA
C.T. CHEN "Irreducible Jordan form realization of a rational matrix"
I.E.E.E. Trans. Aut. Cont. Feb. 1969, p 66-69.
- [14] C.G. GIORGI
A. ISODORI "A new algorithm for the irreducible realization of a rational matrix"
Recherche di Automatica, Vol 11 n° 3 Déc 1971, p 225-239
- [15] Y.L. KUO "On the irreducible Jordan form realization and the degree of rational metrix"
I.E.E.E. Trans. Circ. Theory Vol CT 17 n° 3 Aug 1970 p 322-332.
- [16] P. D'ALESSANDRO
C. GIULIANELLI "A direct procedure for irréducible Jordan form realization"
Recherche di Automatica Vol 2 n° 2 July 1971 p 193-198.
- [17] B.L. HO
R.E. KALMAN "Effective construction of linear state-variable models from input/output function"
Regelungstechnik, Heft 12, 1966, p 545-548.
- [18] E. IRVING "Identification des systèmes"
Note EDF HR 8202/5 EI/MSa-JT Avril 1968
- [19] C. BRUNI
A. ISODORI
A. RUBERTI "A method of realization based on the moments of the impulse response matrix"
I.E.E.E. Trans. Aut. Cont. April 1969, p 203-204.
- [20] L.M. SILVERMAN "Realization of linear dynamical systems"
I.E.E.E. Trans. Aut. Cont. Dec 1971, p 554-567.
- [21] J. RISSANEN "Recursive identification of linear systems"
SIAM J. Control Vol 9 n° 3 Aug. 1971 p 554-567.
- [22] J.E. ACKERMANN
R.S. BUCY "Canonical minimal realization of a matrix of impulse response sequence"
Information and Control Vol 19, 1971, p 224-231
- [23] A.J. TETHER "Construction of minimal linear state variable models from finite input-output data"
I.E.E.E. Trans Aut. Cont. Aug. 1970, p 427-436.
- [24] J.E. ACKERMANN "On partial realization"
I.E.E.E. Trans. Aut. Cont. June 1972, p 381
- [25] C.T. CHEN
D.P. MITAL "A simplified irreducible realization algorithm"
I.E.E.E. Trans. Aut. Cont. Aug. 1972, p 535-537

- [26] D.P. MITAL
C.T. CHEN "Irreducible canonical form réalisation of a rational matrix"
Int. J. Control Vol 18 n° 4, p 881-887, 1973.
- [27] B. GOPINATH "On the identification of linear time invariant systems from input-output data"
The Bell System Tech. Journal Vol 48 n° 5 1969, p 1101 p 1101-1113.
- [28] M.A. BUDIN "Minimale realization of discrete linear systems from input-output observations"
I.E.E.E. Trans. Aut. Cont. Oct. 1971, p 395-401.
- [29] M.A. BUDIN "Minimal realization of continuous linear systems from input-output observations"
I.E.E.E. Trans. Aut. Cont. April 1972, p 252-253.
- [30] Ph DE LARMINAT
A. BARRAUD "Un algorithme de réalisation minimale a partir de données d'entrée-sortie".
C.R.A.S. Paris, t. 275 (24 juillet 1972) p 315-318.
- [31] A. BARRAUD
Ph.DE LARMINAT "Minimal realization and approximation of linear systems from normal operating records"
3rd IFAC Symposium La Haye, 1973, paper TS-5 p 953-956.
- [32] J.E. ACKERMANN "Die minimale Ein - Ausgangs - Beschreibung von mehrgrößensystemen und ihre bestimmung aus Ein-Ausgangs - Messungen"
Regelungstechnik Heft 5, Vol 19 Jahrgang 1971, p 203-206.
- [33] C. BONIVENTO
R. GUIDORZI
G. MARRO "Irreducible canonical réalisation from external data sequences"
Int. J. Control, Vol 17 n° 3, 1973, p 553-563.
- [34] R. GUIDORZI "Canonical structure in the identification of multi-variable système"
3rd IFAC Symposium, La Haye, 1973, paper TS-4, p 943-952.
- [35] S.P. BINGULAC
M. DJORAVIC "Estimation of state variable models using input-output data"
3rd IFAC Symposium, La Haye, 1973, paper TS-14 p 995-1004.
- [36] E. POLAK "An algorithm for reducing a linear, time-invariant différentiel system to state form"
I.E.E.E. Trans Aut. Cont. July 1966, p 577-579
- [37] C.T. CHEN "Irreducibility of dynamical equation realization of sets of differential equations"
I.E.E.E. Trans. Aut. Cont, Fev. 1970, p 131.
- [38] C.J. GUEGUEN "An algebraic algorithm for reducing a differential system to state form"
IFAC Symposium Dusseldorf 1968, paper 1.2.2

Chapitre I

DEFINITIONS ET PROPRIETES DES SYSTEMES DYNAMIQUES LINEAIRES INVARIANTS

o o

o

- 1 - Introduction
 - 2 - Description externe - relation d'entrée-sortie
 - 3 - Description interne - modèle d'état
 - 4 - Réalisation d'une relation d'entrée-sortie
 - 5 - Conclusion
- Bibliographie

Résumé :

Nous nous proposons de préciser, ici, le cadre dans lequel on va se situer : systèmes linéaires invariants de dimension finie, et, à cette occasion, d'introduire une terminologie de base : description externe, interne, réalisation minimale, etc...

A partir de ces éléments, se dégageront les premières définitions concernant les problèmes précis auxquels on s'est attaché : réalisation minimale d'un modèle externe ou d'un ensemble de mesures d'entrées-sorties.

1 - INTRODUCTION

L'étude, la conception et l'automatisation d'un système physique exigent le plus souvent d'en connaître un modèle mathématique. Cette description d'un processus doit être généralement regardée comme un problème d'approximation, dans la mesure où il est rare de pouvoir représenter complètement par un tel modèle tous les éléments contribuant à en décrire son comportement dynamique.

Nous nous sommes limités, ici, aux modèles linéaires stationnaires, décision, a priori, naturelle, face au choix d'une autre structure plus complexe, qu'il faudrait définir, et dont il resterait à justifier le mieux fondé. Elle présente en outre le double avantage de se situer dans un contexte théorique bien défini et de pouvoir déboucher sur des applications pratiques, aussi nombreuses que variées.

Cette modélisation macroscopique, faisant plus ou moins abstraction de la réalité physique du système, nous amène à distinguer principalement parmi les grandeurs le caractérisant : le *temps*, noté t ; les grandeurs *d'entrées*, notées u qui peuvent être des commandes ou des perturbations ; et les grandeurs de *sorties*, notées y , caractérisant ce qui est produit par le système. On peut alors définir un système dynamique comme une transformation dynamique des entrées dans les sorties.

On rappellera successivement deux représentations de cette transformation : soit directement par la relation d'entrée/sortie (§ 2) description externe, soit indirectement en introduisant la notion *d'état*, description interne (§ 3). Les relations entre ces deux descriptions, d'un même système, nous conduiront dans une dernière partie (§ 4) à définir le problème de *réalisation*, thème autour duquel a été développé ce mémoire.

2 - DESCRIPTION EXTERNE : RELATION D'ENTREE/SORTIE

La définition formelle d'un système dynamique d'un point de vue externe est la suivante [1].

Définition I.1 : Un système dynamique Σ est un quintuplet consistant de :

- 1 - Un ensemble des temps T
- 2 - Un ensemble des entrées U
- 3 - Un ensemble des sorties Y
- 4 - Un ensemble Ω des fonctions $T \rightarrow U$, notées $\underline{u}(\cdot)$ et dites *fonctions d'entrées admissibles*. Ω doit : contenir les fonctions constantes -

être fermé pour les translations temporelles - être fermé pour l'enchaînement de fonctions.

5 - Une fonction $\underline{\psi}$ dite *relation d'entrée/sortie* appliquant $\Omega \times T$ dans Y :

$$y(t) = \underline{\psi}(\underline{u}(\cdot), t) \quad (I.1)$$

ne dépendant que des valeurs de $\underline{u}(\cdot)$ antérieures à t (causalité)

Remarque : Si on applique au système une entrée $\underline{u}(\cdot)$ sur un intervalle $[t_1, +\infty]$, notée $\underline{u}[t_1, +\infty]$, il pourra correspondre, à cette entrée, plusieurs sorties $y[t_1, +\infty]$ dépendant de l'entrée appliquée avant t_1 .

Ceci nous conduit à préciser qu'un système est au repos à l'instant t_1 si la sortie $y[t_1, +\infty]$ est définie d'une façon unique par l'entrée $\underline{u}[t_1, +\infty]$; (cette restriction sera précisément levée en introduisant la notion d'état). Nous admettrons par la suite que tout système est au repos pour $t = -\infty$.

Nous nous limitons maintenant aux systèmes dynamiques linéaires stationnaires (invariants) dont le rôle fondamental n'est plus à démontrer tant dans le domaine des servomécanismes que celui, plus récent, de la commande dite "optimale" (optimisation quadratique en particulier).

2.1 - Systèmes continus

Considérons donc un système linéaire stationnaire continu à m entrées et p sorties ; c.a.d. $T = \mathbb{R}$, $U = \mathbb{R}^m$ et $Y = \mathbb{R}^p$. Le système étant linéaire, la fonction $\underline{\psi}$ (I.1) sera, par définition, linéaire en $u(\cdot)$. Cette propriété se conservant pour une combinaison linéaire d'une infinité de fonctions (sous des hypothèses de régularité admises), il suffira pour caractériser le système de connaître la sortie pour une classe d'entrées à partir desquelles toute entrée pourra s'exprimer. Cette classe d'entrée peut se réduire à l'impulsion de Dirac $\delta(t)$ (dans le cas scalaire $m = 1$), puisque toute entrée continue s'écrit :

$$u(t) = \int_{-\infty}^{+\infty} u(\tau) \delta(t-\tau) d\tau \quad (I.2)$$

La réponse du système à l'entrée $\delta(t)$ est, par définition, la *réponse impulsionnelle* du système, notée $g(t)$. La sortie ($p = 1$) pour une entrée quelconque est alors donnée par la relation fondamentale :

$$y(t) = \int_{-\infty}^t g(t-\tau) u(\tau) d\tau \quad (I.3)$$

ou encore :

$$y(t) = \int_0^{\infty} g(\tau) u(t-\tau) d\tau \quad (I.4)$$

notée indifféremment :

$$y(t) = g * u \quad (I.5)$$

* symbolisant le produit de convolution (continu)

Remarque : On s'est placé ici dans l'hypothèse habituelle où l'impulsion $\delta(t)$ est appliquée à l'instant initial $t_0 = 0$. Plus généralement, $g(t)$ est la réponse à l'entrée $\delta(t-t_0)$, qui par causalité est nulle pour $t < t_0$.

Par ailleurs, l'équation (I.5) s'étend aisément aux systèmes multi-variables ; en notant G , matrice $p \times m$, d'éléments $g_{ij}(t)$ (réponse impulsionnelle de la sortie i relative à l'entrée j), il vient :

$$\underline{y}(t) = G * \underline{u} \quad (I.6)$$

Introduisons maintenant la transformation de Laplace, notée $\mathcal{L}(\cdot)$. Admettons que $\underline{u}(t)$ et $\underline{y}(t)$ soient identiquement nuls pour $t \leq t_0 = 0$, de sorte que le système est implicitement au repos, à l'instant t_0 . Définissons alors les quantités suivantes :

$$\begin{aligned} \underline{u}(s) &= \mathcal{L}[\underline{u}(t)] \\ \underline{y}(s) &= \mathcal{L}[\underline{y}(t)] \\ G(s) &= \mathcal{L}[G(t)] \end{aligned} \quad (I.7)$$

où $G(s)$ est ce qu'on appelle la *matrice de transfert* du système. La relation fondamentale (I.6) devient alors :

$$\underline{y}(s) = G(s) \cdot \underline{u}(s) \quad (I.8)$$

2.2 - Systèmes discrets

La différence essentielle avec le cas continu se situe dans la nature de T qui est pris, ici, pour simplifier, identique à Z , l'ensemble des entiers relatifs. En suivant une démarche analogue avec celle du paragraphe précédent, il convient d'introduire le symbole de Kronecker δ_t , à l'aide duquel toute entrée discrète $u_t^\dagger$ ($m = 1$) peut être exprimée.

$$u_t = \sum_{\tau=-\infty}^{+\infty} u_\tau \cdot f_{t-\tau} \quad (I.9)$$

On définira alors la réponse impulsionnelle discrète f_t (ou *séquence de pondération*) comme étant la réponse du système à l'entrée δ_t . De même, la sortie ($p = 1$), pour une entrée quelconque, est donnée par :

$$y_t = \sum_{\tau=-\infty}^t f_{t-\tau} \cdot u_\tau \quad (I.10)$$

[†] On note l'entrée à l'instant t : u_t et non $u(t)$ pour la distinguer du signal continu.

ou encore :

$$y_t = \sum_{\tau=0}^{\infty} f_{\tau} u_{t-\tau} \quad (I.11)$$

noté également :

$$y_t = f \star u \quad (I.12)$$

★ désignant le produit de convolution (discret)

On a bien sûr $f_t \equiv 0$ pour $t < 0$ et dans le cas multivariable (I.12) devient :

$$\underline{y}_t = F \star \underline{u} \quad (I.13)$$

où F est la matrice $p \times m$ des séquences de pondération.

Cette dernière relation s'exprime encore en prenant la transformée en z :

$$\underline{y}(z) = F(z) \cdot \underline{u}(z) \quad (I.14)$$

où $F(z)$ est la matrice de transfert (discrète) du système.

Remarque : On suppose implicitement que $\underline{u}_t$ et $\underline{y}_t$ sont nuls pour $t < 0$.

3 - DESCRIPTION INTERNE : MODELE D'ETAT -

On a vu que la relation d'entrée/sortie, $\underline{\psi}$, ne permettait de déterminer la sortie du système à l'instant t , que si l'on connaissait tout le passé de l'entrée. On conçoit facilement les difficultés d'utiliser parfois cette description si le système n'est pas au repos à l'instant initial considéré. Ceci nous amène à introduire le concept d'état que l'on définit comme étant un ensemble de grandeurs permettant de résumer le passé en vue de l'étude de l'avenir. En d'autres termes, l'état représente l'ensemble des conditions initiales qui conjointement à la connaissance de l'entrée $\underline{u} [t_0, \infty]$ permet de caractériser d'une façon unique la sortie $\underline{y} [t_0, +\infty]$. Nous complèterons cet aspect qualitatif du problème par la définition formelle suivante [1] :

Définition I.2 : Un système dynamique Σ est un septuplet consistant des points 1-2-3 et 4 de la définition I.1 et de :

5 - Un ensemble d'état X

6 - Une fonction de transition d'état $\underline{\phi}$, appliquant $T \times T \times X \times \Omega$ dans X :

$$\underline{x}(t) = \underline{\phi}(t, t_0, \underline{x}(t_0), \underline{u}[t_0, t]) \quad \forall t > t_0 \quad (I.15)$$

qui exprime que l'on passe de l'état $\underline{x}(t_0)$ à $\underline{x}(t)$ en appliquant $\underline{u}[t_0, t]$ (causalité).

La fonction ϕ doit être :

$$\text{cohérente : } \phi(t; t; \underline{x}(t); \underline{u}[t, t]) = \underline{x}(t) ; \forall t, \underline{x}(t) \text{ et } \underline{u}(.) \quad (\text{I.16})$$

$$\text{composable : } \phi(t_2; t_0; \underline{x}(t_0); \underline{u}[t_0, t_2]) = \phi(t_2; t_1; \phi(t_1; t_0; \underline{x}(t_0); \underline{u}[t_0, t_1]); \underline{u}[t_1, t_2]) \quad (\text{I.17})$$

$$\forall t_0, t_1, t_2 : t_0 < t_1 < t_2$$

7 - Une fonction de sortie $\underline{y}$ appliquant $X \times U \times T$ dans Y :

$$\underline{y}(t) = \underline{y}(\underline{x}(t), \underline{u}(t), t) ; \quad (\text{I.18})$$

définissant la sortie en fonction de l'entrée et de l'état à l'instant t .

Nous développons rapidement dans les deux paragraphes suivants la classe des systèmes constituée par les systèmes différentiels (aux différences) linéaires invariants de dimension finie dont la fonction ϕ est directement définie comme l'intégrale (la solution) générale d'une équation différentielle (aux différences) dite *équation d'état*.

3.1 - Systèmes continus

Un système dynamique linéaire stationnaire continu, de dimension finie ($X = \mathbb{R}^n$) est caractérisé par les équations :

$$\begin{cases} \dot{\underline{x}} = A \underline{x} + B \underline{u} \\ \underline{y} = C \underline{x} + D \underline{u}^+ \end{cases} \quad (\text{I.19})$$

où A , B , C et D sont des matrices constantes de dimensions appropriées.

En intégrant l'équation différentielle et en prenant $t_0 = 0$ comme instant initial, il vient :

$$\underline{y}(t) = C e^{At} \underline{x}(0) + \int_0^t C e^{A(t-\tau)} B \underline{u}(\tau) d\tau + D \underline{u}(t) \quad (\text{I.20})$$

Supposons le système initialement au repos ($\underline{x}(0) = \underline{0}$) et appliquons lui successivement les entrées $\underline{u}(t) = \underline{0}$ sauf la $i^{\text{ème}}$ composante égale à $\delta(t)$, on en déduit immédiatement la réponse impulsionnelle de (I.19), soit :

$$G(t) = C e^{At} B + D \delta(t) \quad (\text{I.21})$$

Prenons-en la transformée de Laplace, il vient :

$$G(s) = C(s\mathbf{1} - A)^{-1} B + D \quad (\text{I.22})$$

+ Une forme plus générale pourrait être obtenue en ajoutant des termes en $\underline{u}'$, $\underline{u}''$...

3.2 - Systèmes discrets

Par analogie avec (I.19), un système dynamique, linéaire discret, de dimension finie, est caractérisé par les équations :

$$\begin{cases} \underline{x}_{t+1} = \Phi \underline{x}_t + \Gamma \underline{u}_t \\ \underline{y}_t = H \underline{x}_t + D \underline{u}_t \end{cases} \quad (I.23)$$

Un calcul immédiat conduit à l'expression :

$$\underline{y}_t = H \Phi^t \underline{x}_0 + \sum_{i=0}^{t-1} H \Phi^{t-i-1} \Gamma \underline{u}_i + D \underline{u}_t \quad (I.24)$$

contre partie discrète de (I.20). Toujours par analogie avec le cas continu, on en déduit la relation donnant la séquence de pondération du système décrit par (I.23).

$$\begin{cases} F_0 = D \\ F_t = H \Phi^{t-1} \Gamma \quad t > 0 \end{cases} \quad t \in \mathbb{N}^+ \quad (I.25)$$

Enfin, par transformation en z de (I.25), il vient :

$$F(z) = H(zI - \Phi)^{-1} \Gamma + D \quad (I.26)$$

4 - REALISATION D'UNE RELATION D'ENTREE/SORTIE

On vient de voir à travers les équations (I.21) - (I.22), d'une part, et (I.25) - (I.26), d'autre part, qu'il est relativement facile de donner la description externe d'un système linéaire représenté par sa description interne.

Le problème inverse, appelé problème de *réalisation* d'une relation d'entrée/sortie est beaucoup moins simple.

Définition I.3 : Etant donné $G(s)$ (ou $G(t)$), les matrices (A, B, C, D) définissent une *réalisation* de G , si, et seulement si, la relation (I.22) (ou (I.21)) est satisfaite.

La version discrète du problème s'énonce de la même manière :

Définition I.4 : Etant donné $F(z)$ (ou F_t), les matrices (Φ, Γ, H, D) définissent une *réalisation* de F , si, et seulement si, la relation (I.26) (ou (I.25)) est satisfaite.

D'une façon générale :

Définition I.5 : On dit qu'une réalisation est *minimale* ou *irréductible*, s'il n'existe pas d'autres réalisations dont la dimension de X soit plus petite.

On peut dès lors se poser trois questions :

- Existe-t-il une solution à ce problème ? (réalisabilité)
- Si oui, est-elle unique ? (classe d'équivalence)
- Comment en déterminer une solution ? (algorithmes de réalisation)

Nous répondons maintenant aux deux premiers points, le troisième sera traité au chapitre IV.

4.1 - Réalisabilité

Le problème posé consiste à savoir si, étant donné une relation d'entrée/sortie ($G(s)$, $G(t)$, $F(z)$ ou $F(t)$), il existe a priori un modèle d'état de la forme (I.19) (ou (I.23)), de dimension finie, ayant même relation d'entrée/sortie.

Introduisons d'abord la ...

Définition I.6 : Une matrice $G(s)$ dont les éléments sont des fonctions rationnelles en s est dite *propre* si $G(\infty)$ est une matrice constante ; $G(s)$ est *strictement propre* si $G(\infty) = 0$.

Exemple : $G(s) = C(sI - A)^{-1}B + D$ est une matrice propre ; $G(\infty) = D$.

Si $D = 0$, $G(s)$ est alors strictement propre.

... ce qui nous permet de rappeler le résultat suivant [2] :

Théorème I.1 : Une matrice de transfert $G(s)$ est *réalisable* par un système dynamique fini de la forme (I.19), si, et seulement si, $G(s)$ est une matrice rationnelle propre.

Remarque : Ce théorème s'applique ipso facto aux systèmes discrets en remplaçant $G(s)$ par $F(z)$.

Corollaire I.1 : La matrice $G(t)$ est *réalisable* par un système dynamique fini de la forme (I.19) si, et seulement si, $g_{ij}(t)$ est une combinaison linéaire de termes $t^k e^{\lambda_\ell t}$.

Remarque : Pour les systèmes discrets, f_t^{ij} doit être une combinaison linéaire de termes $t^k \alpha_\ell^t$.

4.2 - Classe d'équivalence

Soit un système dynamique $\Sigma(A, B, C, D)$, peut-on déterminer un autre système de même dimension $\tilde{\Sigma}(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ ayant même relation d'entrée/sortie ? Les éléments de la réponse sont donnés par la ...

Définition I.1 : Deux systèmes dynamiques de même dimension, $\Sigma(A, B, C, D)$ et $\tilde{\Sigma}(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ sont *équivalents*, s'ils sont reliés par la relation : $\tilde{\underline{x}} = P \underline{x}$

où P est une matrice constante régulière.

... et le ...

Théorème I.2 [2] : Deux systèmes équivalents Σ et $\tilde{\Sigma}$ ont même relation d'entrée/sortie.

De la définition I.7, il vient :

$$\tilde{A} = P A P^{-1} ; \quad \tilde{B} = P B ; \quad \tilde{C} = C P^{-1} \quad \text{et} \quad \tilde{D} = D \quad (\text{I.27})$$

Quant au théorème I.2, il implique :

$$\left\{ \begin{array}{l} G(t) = C e^{At} B + D \delta(t) = \tilde{C} e^{\tilde{A}t} \tilde{B} + \tilde{D} \delta(t) \\ G(s) = C(sI-A)^{-1} B + D = \tilde{C}(sI-\tilde{A})^{-1} \tilde{B} + \tilde{D} \\ F_t = H \Phi^{t-1} \Gamma = \tilde{H} \tilde{\Phi}^{t-1} \tilde{\Gamma}, \quad t > 0 \in \mathbb{N} ; \quad D = \tilde{D} = F_0 \\ F(z) = H(zI-\Phi)^{-1} \Gamma + D = \tilde{H}(zI-\tilde{\Phi})^{-1} \tilde{\Gamma} + \tilde{D} \end{array} \right. \quad (\text{I.28})$$

Remarque : La réciproque du théorème I.2 est généralement fausse.

Deux réalisations de même dimension ne sont pas, en effet, nécessairement équivalentes, sauf s'il s'agit de réalisations minimales. Ce point sera repris à l'occasion de l'étude de la structure de l'espace d'état, chapitre II.

En conséquence, s'il existe une solution au problème de réalisation, il en existe plusieurs infinités de dimensions éventuellement différentes. En particulier, une réalisation minimale n'est unique que modulo un changement de base.

4.3 - Généralisation du concept de réalisation

Le problème de réalisation consiste, on vient de le voir, à déterminer une description interne d'un système dynamique connu par sa relation d'entrée/sortie (ψ). Elle a été caractérisée, jusqu'ici, soit par la matrice de transfert en s (ou en z), soit par la matrice des réponses impulsionnelles (discrètes). Nous avons signalé, dans l'introduction, une autre caractérisation qui est celle donnée par l'opérateur différentiel (aux différences), le plus souvent de la forme : [3]

$$L \left(\frac{d}{dt} \right) \underline{y}(t) = M \left(\frac{d}{dt} \right) \underline{u}(t) \quad (\text{I.29})$$

$$\text{ou} \quad L(z) \cdot \underline{y}_t = M(z) \cdot \underline{u}_t$$

les quantités L et M étant des matrices polynomiales de dimension $p \times p$ et $p \times m$.

† z est ici considéré comme l'opérateur avance

Une généralisation, qui nous paraît beaucoup plus intéressante, est celle qui consiste à définir (en quelque sorte) $\underline{\psi}$ directement par les données $\underline{u}[t_0, t_1]$, $\underline{y}[t_0, t_1]$.

Ceci suppose implicitement que l'entrée $\underline{u}(\cdot)$ soit suffisamment riche pour que la sortie $\underline{y}(\cdot)$ soit représentative du spectre du système. Ce problème, *identifiabilité*, sera développé au chapitre IV au moment d'aborder les algorithmes de réalisation correspondants.

Pour l'instant, nous précisons cette autre approche par la ...

Définition I.8 : La réalisation d'un ensemble de mesures $\underline{u}[t_0, t_1]$, $\underline{y}[t_0, t_1]$ est un modèle d'état de la forme (I.19) ou (I.23), satisfaisant les données, pour un certain état (initial) $\underline{x}$ à l'instant t_0 .

Remarque : Une réalisation de dimension minimale, compatible avec les données ne sera pas ici nécessairement irréductible au sens introduit précédemment, à moins que le système considéré ne soit par exemple au repos à l'instant t_0 . En d'autres termes, l'identité sur l'intervalle $[t_0, t_1]$ des sorties données et modèle n'impliquera pas toujours celle des matrices de transfert..

Cette contradiction (apparente) qui s'éclaircira d'elle-même au cours du chapitre II est ce qui différencie ici ce problème spécifique de réalisation (Définition I.8) des précédents.

5 - CONCLUSION

Dans ce premier chapitre, on a ainsi défini le cadre de notre mémoire : systèmes linéaires stationnaires de dimension finie, et introduit, au moins d'une manière qualitative, les premiers éléments caractéristiques du problème considéré : réalisations minimales.

L'importance que revêt déjà intuitivement le concept d'irréductibilité s'exprimera en termes précis à l'issue de l'analyse des propriétés et de la structure de l'espace d'état X que l'on aborde maintenant.

BIBLIOGRAPHIE

- [1] R.E. KALMAN "Topics in mathematical system theory"
P.L. FALB Mc Graw Hill Book Company 1969.
M.A. ARBIB

- [2] C.T. CHEN "Introduction to linear system theory"
1970 Halt, Rinehart and Winston, Inc.

- [3] A. FOSSARD "Commande des systèmes multidimensionnels"
Dunod 1972

Chapitre II

STRUCTURE DE L'ESPACE D'ETAT

o o
o

- 1 - Introduction
- 2 - Préliminaires
- 3 - Commandabilité et accessibilité
- 4 - Observabilité et restructurabilité
- 5 - Structure de l'espace d'état définie
par l'accessibilité et l'observabilité
- 6 - Réalisation minimale d'un système dynamique
- 7 - Conclusion
- Bibliographie

Résumé :

On rappelle, dans une première partie de ce chapitre, les concepts fondamentaux caractérisant l'espace d'état. A partir de ces développements, effectués parallèlement dans le cas continu et discret, on fait apparaître la décomposition "accessibilité-observabilité" de cet espace, sur laquelle repose la notion d'irréductibilité des réalisations.

Dans un but d'unification, on s'est attaché à exposer dans cette deuxième partie des résultats indépendants de la nature de l'ensemble T (le temps).

1 - INTRODUCTION

Parmi la multiplicité des solutions relatives au problème de réalisation, on a mis en évidence une classe particulière, dite minimale. Les autres, de dimension (par définition) supérieure confèrent donc au modèle d'état correspondant une complexité non significative vis-à-vis du comportement entrée/sortie du système.

Derrière cet aspect qualitatif du problème apparaît en filigrane les deux concepts classiques de commandabilité et d'observabilité. Alors qu'il est notoire que ces notions caractérisent sans ambiguïté la structure des systèmes linéaires stationnaires continus, on relève souvent dans la littérature, pour le cas discret, des imprécisions de vocabulaire, et parfois même des erreurs. Ce paradoxe apparent provient du fait que la spécification complète de l'espace d'état nécessite, en toute rigueur, l'introduction de deux autres concepts (dus à Kalman) : l'accessibilité et la restructurabilité.

De ces considérations développées § 3 et 4, il ressort que seules les parties accessibles et observables de l'espace d'état ont un lien avec une relation d'entrée/sortie du système (quelle que soit la nature de T).

Ce résultat nous permet, après avoir rappelé (§ 5) la décomposition canonique correspondante de l'espace d'état, de caractériser, en termes précis, l'irréductibilité d'une réalisation. A cette occasion, nous emprunterons une terminologie due initialement à Zeiger, faisant abstraction de la nature discrète ou continue du système (§ 6).

C'est également dans un but d'unification que l'on fera appel au cours de ce chapitre à quelques résultats d'algèbre linéaire, rappelés dans un paragraphe préliminaire.

2 - PRELIMINAIRES

Considérons l'application linéaire de R^n dans R^m définie par la relation :

$$A \underline{x} = \underline{b} \quad (II.1)$$

où A est une matrice $m \times n \in R^m \times R^n$ et où $\underline{x}$ et $\underline{b}$ sont deux éléments de R^n et R^m respectivement. Définissons par $R(A)$ l'espace engendré par les colonnes de A . On notera $\rho(A)$ la dimension de cet espace, qui n'est autre que le rang de la matrice A , par suite on a :

$$\rho(A) \leq \min (m,n) \quad (II.2)$$

Soit $N(A)$ le noyau de A (ensemble des $\underline{x} \in \mathbb{R}^n : A \underline{x} = \underline{0}$) et $v(A)$ sa dimension. Rappelons alors que :

Proposition II.1 [1] : $\rho(A) + v(A) = n$ (II.3)

Proposition II.2 [1] : Etant donné $\underline{b} \in \mathbb{R}^m$, il existe une solution $\underline{x} \in \mathbb{R}^n$, vérifiant (II.1), si, et seulement si, $\underline{b} \in R(A)$.

Proposition II.3 [1] : Etant donné $A \in \mathbb{R}^m \times \mathbb{R}^n$ et pour tout $\underline{b} \in \mathbb{R}^m$, il existe une solution $\underline{x} \in \mathbb{R}^n$, vérifiant (II.1), si, et seulement si, $\rho(A) = m$.

Remarque : Ceci n'est donc possible que si $m \leq n$ d'après (II.2). Par ailleurs, la solution est unique si, et seulement si, $v(A) = 0$, alors $m = n$, de sorte que A est régulière.

Proposition II.4 : L'application définie par A est surjective si, et seulement si, $\rho(A) = m$.

Remarque : On a alors nécessairement $m \leq n$ d'après (II.2).

Démonstration : Elle découle de la définition de la surjection (existence d'au moins une solution $\underline{x}$ pour tout $\underline{b}$) et de la proposition II.3.

Proposition II.5 : L'application définie par A est injective si, et seulement si, $\rho(A) = n$.

Remarque : On a alors nécessairement $m \geq n$, d'après (II.2).

Démonstration : S'il existe une solution $\underline{x}$, elle est unique si, et seulement si $v(A) = 0 \iff \rho(A) = n$ d'après (II.2). Par suite, $\rho(A) = n$ est condition nécessaire et suffisante pour qu'il existe *au plus* une solution à (II.1).

On rappellera enfin deux autres résultats, concernant le rang de certaines matrices, résultats qui interviendront fréquemment dans la suite. Soit donc X et Y deux matrices respectivement de dimension $n \times n$ et $n \times m$ et définissons par Z_j la matrice :

$$Z_j = [Y, XY, X^2Y, \dots, X^jY] \quad (II.4)$$

Proposition II.6 [1] : Si j est le plus petit entier tel que $\rho(Z_j) = \rho(Z_{j+1})$ alors $\rho(Z_k) = \rho(Z_j)$ pour tout $k > j$.

Proposition II.7 [1] : En notant $r = \rho(Y)$ et $\bar{n} (\leq n)$ le degré du polynome minimal de la matrice X , on a :

$$j \leq \min (n-r, \bar{n}-1) \quad (II.5)$$

où j est défini dans la proposition II.6.

3 - COMMANDABILITE ET ACCESSIBILITE

Les ensembles U , X et T , qui interviennent ici, sont ceux précédemment définis au chapitre I. Σ , $\Sigma(A,B,C,D)$ ou $\Sigma(\Phi, \Gamma, H, D)$ désigneront un système dynamique linéaire stationnaire de dimension n (celle de X) sauf indication contraire.

On appellera enfin *évènement* le couple ordonné $(\underline{x}, \tau)$, $\in X \times T$.

Définition II.1 [2], [3]

Un évènement $(\underline{x}, \tau)$ est *commandable* (à l'état nul), si, et seulement si, il existe un instant $t \geq \tau$ fini et une entrée $\underline{u}[\tau, t]$ qui amène $(\underline{x}, \tau)$ en $(\underline{0}, t)$.

Définition II.2 [2][†]

Un évènement $(\underline{x}, \tau)$ est *accessible* (depuis l'état nul), si et seulement si, il existe un instant $s \leq \tau$ fini et une entrée $\underline{u}[s, \tau]$ qui amène $(\underline{0}, s)$ en $(\underline{x}, \tau)$.

Pour préciser, on parlera de *complète* commandabilité (accessibilité) à l'instant τ si, et seulement si, tout évènement $(\underline{x}, \tau)$, τ fixé $\forall \underline{x} \in X$, est commandable (accessible). Par ailleurs, pour les systèmes stationnaires où l'instant τ n'intervient plus (seulement $t-\tau$ ou $\tau-s$), on dira plus simplement, par abus de langage, que le système Σ est commandable (accessible).

Remarques :

- La commandabilité a été définie de telle sorte qu'elle représente une condition naturelle, nécessaire et suffisante, pour pouvoir résoudre le problème du régulateur. Rappelons que celui-ci consiste à ramener un état initial quelconque à un autre état donné, généralement nul. Si on impose explicitement certaines contraintes sur le temps, les variations de $\underline{u}$ ou/et de $\underline{x}$, on fait alors référence à un problème de commande.

- A propos de l'accessibilité, contre partie de la commandabilité pour un temps s'écoulant à rebours (fig. 1), elle va intervenir comme une des caractéristiques fondamentales des problèmes de réalisation minimale, cf. §3.3.

† Dans [4], cette définition est prise comme étant celle de la commandabilité ; ceci évite d'emblée toutes difficultés avec le cas discret, telles qu'elles seront soulignées par les corollaires (II.6) et (II.7) au § 3.2.

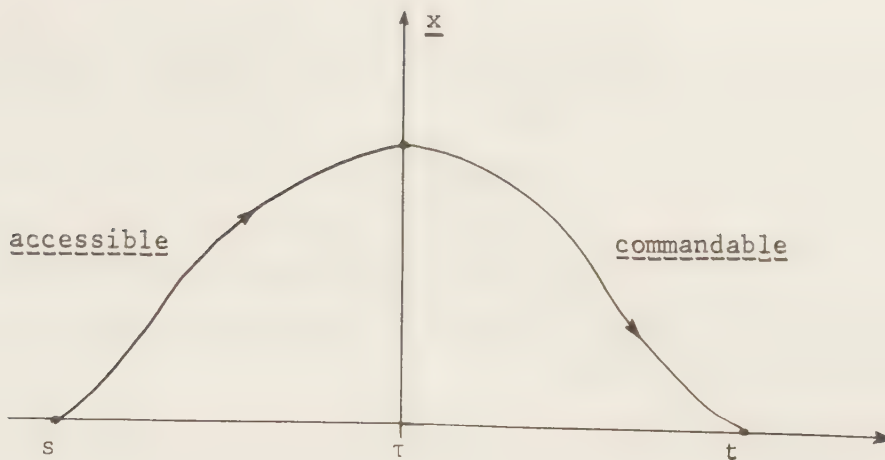


Fig. 1

La question que l'on peut se poser maintenant est de savoir quand un système dynamique Σ est, ou n'est pas, commandable ou accessible. Pour cela, exprimons l'évolution de l'état du système en fonction du temps. On a :

$$\dot{\underline{x}} = A \underline{x} + B \underline{u} \quad (\text{II.6})$$

ou

$$\underline{x}_{t+1} = \Phi \underline{x}_t + \Gamma \underline{u}_t \quad (\text{II.7})$$

suivant la nature de T . Par suite, deux événements $(\underline{x}, t_1)$ et $(\underline{x}, t_2)$, $t_2 \geq t_1$ sont reliés par les équations :

$$\underline{x}(t_2) = e^{A(t_2-t_1)} \underline{x}(t_1) + \int_{t_1}^{t_2} e^{A(t_2-t)} B \underline{u}(t) dt \quad (\text{II.8})$$

ou

$$\underline{x}_{t_2} = \Phi^{(t_2-t_1)} \underline{x}_{t_1} + \sum_{t=t_1}^{t_2-1} \Phi^{t_2-t-1} \Gamma \underline{u}_t \quad (\text{II.9})$$

3.1 - Critère d'accessibilité

Reprenons pour cela les expressions (II.8) et (II.9) en posant $t_1 = 0$ (ce qui n'est nullement restrictif pour les systèmes stationnaires) et notons $t_2 = t$. Le problème d'accessibilité implique que l'état initial soit nul. Par suite, (II.8) s'écrit :

$$\underline{x}(t) = \int_0^t e^{A(t-\tau)} B \underline{u}(\tau) d\tau \quad (\text{II.10})$$

et (II.9) devient :

$$\underline{x}_t = \sum_{\tau=0}^{t-1} \Phi^{t-\tau-1} \Gamma \underline{u}_\tau \quad (\text{II.11})$$

Il est clair que (II.10) et (II.11) définissent une application linéaire de U dans X , que l'on note :

$$U \xrightarrow{g} X \quad (II.12)$$

Par définition Σ est accessible si, et seulement si, étant donné l'évènement $(\underline{x}, t)$, il existe *au moins* une entrée $\underline{u}(\cdot) \in U$ vérifiant (II.10) ou (II.11). Ceci signifie que l'application g doit être surjective. Formalisons (cf. aussi [2]).

Théorème II.1 : Le système dynamique Σ est accessible si, et seulement si, l'application $U \xrightarrow{g} X$ est surjective.

Ce résultat est important à plusieurs titres. Sur un plan théorique, cette condition pourra intervenir directement comme élément caractéristique d'une réalisation minimale, indépendamment de la nature exacte de T , cf § 5. Par ailleurs, elle permettra de retrouver par des raisonnements simples certains critères classiques, sous forme matricielle.

- Explicitons d'abord le cas discret. De (II.11), il vient :

$$\underline{x}_t = \Delta_t(\Phi, \Gamma) \underline{u}' \quad (II.13)$$

en notant :

$$\Delta_t(\Phi, \Gamma) \triangleq [\Gamma, \Phi\Gamma, \dots, \Phi^{t-1}\Gamma] \begin{matrix} \updownarrow n \\ \longleftarrow t \times m \end{matrix} \quad \text{et} \quad \underline{u}' = \begin{matrix} \xleftarrow{1} \\ \left[\begin{matrix} \underline{u}_{t-1} \\ \vdots \\ \underline{u}_0 \end{matrix} \right] \updownarrow t \times m \end{matrix}$$

D'après la proposition II.4, une condition nécessaire et suffisante pour que g soit surjective est que $\rho(\Delta_t) = n$, c'est-à-dire que Δ_t soit de rang maximum. Enfin, compte tenu de la proposition II.7, on sait que si un tel t existe, celui-ci est borné supérieurement par $\bar{t} = \min(n-r+1, \bar{n})$ où $r = \rho(\Gamma)$ et $\bar{n}$ est le degré du polynôme minimal de Φ . On aura donc :

Corollaire II.1 : L'application g définie par (II.11) est surjective si, et seulement si, $\rho(\Delta_{\bar{t}}) = n$.

On en déduit immédiatement le ...

Corollaire II.2 : Le système $\Sigma(\Phi, \Gamma, H, D)$ est accessible si, et seulement si, $\rho(\Delta_{\bar{t}}) = n$.

Remarque : L'accessibilité ne dépendant que de Φ et Γ , il paraît naturel de parler de l'accessibilité de la paire (Φ, Γ) , en se référant à celle du système Σ correspondant.

- Le cas continu se traite d'une manière analogue. Rappelons à cet effet que e^{At} peut s'exprimer à l'aide d'une somme finie de puissances de A, soit :

$$e^{At} = \sum_{k=0}^{\bar{n}-1} \alpha_k(t) A^k \quad (\text{II.14})$$

où $\bar{n}$ est le degré du polynôme minimal de A.

Dans ces conditions, l'équation (II.10) s'écrit maintenant :

$$\underline{x}(t) = [B, AB, \dots, A^{\bar{n}-1}B] \begin{bmatrix} \int_0^t \alpha_0(t-\tau) u(\tau) d\tau \\ \vdots \\ \int_0^t \alpha_{\bar{n}-1}(t-\tau) u(\tau) d\tau \end{bmatrix} \quad (\text{II.15})$$

Par suite, l'existence d'au moins une solution $\underline{u}[0,t]$ étant donné $\underline{x}(t)$, c'est-à-dire g surjective, se traduit, d'après la proposition (II.4) par le ...

Corollaire II.3 : L'application g définie par (II.10) est surjective, si, et seulement si, $\rho[\Delta_{\bar{t}}^-(A,B)] = n$.

Remarque : On a noté une fois de plus $\bar{t} = \min(n-r, \bar{n}-1)$ avec $r = \rho(B)$.

Compte tenu de ce qui précède on a, bien sûr, le ...

Corollaire II.4 : La paire (A,B) est accessible si, et seulement si, $\rho[\Delta_{\bar{t}}^-(A,B)] = n$.

Il est particulièrement intéressant de noter la similitude entre les corollaires (II.2) et (II.4) alors que la signification des paires (A,B) et (Φ, Γ) est totalement différente. Cette remarque suggère de voir les propriétés des systèmes dynamiques essentiellement liées à celles de formes purement algébriques indépendamment de la nature discrète ou continue du temps. C'est en fait rigoureux, et la justification complète de ce point de vue a été donnée par Kalman, via la théorie des modules [2]. Notre propos n'est pas d'aborder de telles considérations, mais de mettre en évidence des moyens simples permettant de traiter parallèlement systèmes continus et discrets sans avoir à les distinguer dans l'énoncé des résultats fondamentaux ; le théorème II.1 en est un exemple.

C'est dans cet esprit que l'on aborde maintenant la commandabilité des systèmes.

3.2 - Critère de commandabilité

Revenons aux relations (II.8) et (II.9) en posant toujours $t_2 = t$ et $t_1 = 0$ mais en supposant cette fois que l'état final est nul. De (II.8), il vient donc :

$$\underline{0} = e^{At} \underline{x}(0) + \int_0^t e^{A(t-\tau)} B \underline{u}(\tau) d\tau$$

que l'on écrit sous la forme :

$$M \underline{x}(0) = \int_0^t e^{A(t-\tau)} B \underline{u}(\tau) d\tau \quad (\text{II.16})$$

en notant ici $M = -e^{At}$.

En opérant de la même manière avec (II.9), on obtient :

$$M \underline{x}_0 = \sum_{\tau=0}^{t-1} \phi^{t-\tau-1} \Gamma \underline{u}_\tau \quad (\text{II.17})$$

avec maintenant $M = -\phi^t$.

Les relations (II.16) et (II.17) définissent encore une application linéaire de U dans un certain espace $X' \subseteq X$ (au sens large), que l'on note :

$$U \xrightarrow{g'} X' \quad (\text{II.18})$$

Remarque : L'espace X' n'est autre que $R(M)$, l'ensemble $M\underline{x}$, $\forall \underline{x} \in X$.

En se référant à la définition de la commandabilité, on en déduit le ...

Théorème II.2 : Le système dynamique Σ est commandable si, et seulement si, l'application g' est surjective.

Pour poursuivre, il est maintenant nécessaire de distinguer deux cas : M est régulière alors $X' \equiv X$, ou M est singulière et $X' \subset X$ (strictement).

Il est clair que si $X' \equiv X$, la commandabilité et l'accessibilité s'impliquent mutuellement. C'est en particulier le cas pour les systèmes continus, puisque $M = e^{At}$ et $M^{-1} = e^{-At}$; par suite nous avons le ...

Corollaire II.5 : La paire (A,B) est commandable si, et seulement si, $\rho[\Delta_t^-(A,B)] = n$.

Dans le cas discret les deux éventualités sont possibles selon que ϕ est régulière ou non. D'après ce qui précède, on a :

Corollaire II.6 : La paire (ϕ, Γ) , ϕ régulière, est commandable si, et seulement si, $\rho[\Delta_t^-(\phi, \Gamma)] = n$.

Supposons maintenant que Φ soit singulière, et reprenons les termes du théorème II.2. D'après la proposition II.2, il s'agit que pour tout $\underline{x}_0 \in X$, il existe t et $\underline{u}(\cdot)$ tels que, d'après (II.17) et (II.13) :

$$- \Phi^t \underline{x}_0 = \Delta_t(\Phi, \Gamma) \underline{u}' \quad (\text{II.19})$$

$$\iff R(\Phi^t) \subseteq R[\Delta_t(\Phi, \Gamma)] \quad (\text{II.20})$$

Compte tenu de la proposition (II.7), on énoncera ce dernier résultat sous forme du ...

Corollaire II.7 : La paire (Φ, Γ) , Φ singulière, est commandable, si et seulement si, $R[\Phi^t] \subseteq R[\Delta_t(\Phi, \Gamma)]$ pour un $t \leq \bar{t}^+$

Remarques :

- Lorsqu'un système Σ est à la fois commandable et accessible, il s'en suit que tout événement $(\underline{x}, t_1)$ peut être transformé, en un temps fini, dans le futur, en un événement $(\underline{x}, t_2)$ donné. C'est en fait toujours le cas si l'on fait exception des systèmes discrets avec Φ singulière. Ceci explique et justifie peut être que l'on prenne souvent cette dernière propriété comme définition de la commandabilité sans introduire la notion d'accessibilité, [1], [5], [6].

- Cette exception que l'on vient de signaler, a priori peu naturelle, s'explique aisément si l'on considère que l'hypothèse Φ singulière correspond à une situation particulière (système à retard pur) qui n'a pas d'équivalent pour un système différentiel continu de dimension finie.

- En toute généralité, l'ensemble des états accessibles est inclus dans l'ensemble des états commandables. Autrement dit :

$$\begin{aligned} \Sigma \text{ accessible} &\Rightarrow \Sigma \text{ commandable} \\ \Sigma \text{ commandable} &\not\Rightarrow \Sigma \text{ accessible} \end{aligned} \quad (\text{II.21})$$

Exemple :

Considérons le système :

$$\underline{x}_{t+1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \underline{x}_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_t$$

On a $\bar{t} = 1$, d'où :

$$\Delta_{\bar{t}} = [\Gamma, \Phi \bar{t}] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

† Si Φ est régulière, les corollaires II.6 et II.7 sont alors identiques, voir aussi Dorato [9].

En appliquant le corollaire II.2, on en déduit :

$$\rho \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 1 \Rightarrow \Sigma \text{ non accessible}$$

Par ailleurs, d'après le corollaire II.7 :

$$R[\Phi] = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \subseteq R \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \Sigma \text{ est commandable}$$

3.3 - Décomposition de l'espace d'état (lère partie)

Pour clarifier l'exposé, on adopte les conventions suivantes :

$\mathcal{A}$ désigne le sous-espace de X , accessible

$\bar{\mathcal{A}}$ désigne le sous-espace de X , non-accessible

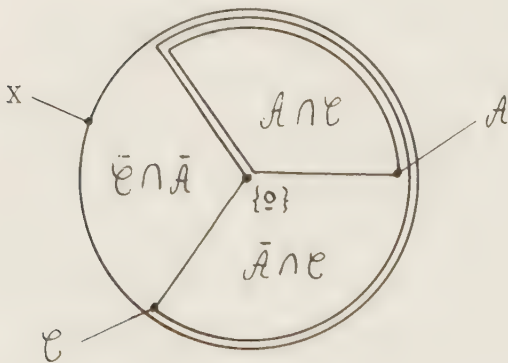
$\mathcal{C}$ désigne le sous-espace de X , commandable

$\bar{\mathcal{C}}$ désigne le sous-espace de X , non-commandable

... de sorte que ...

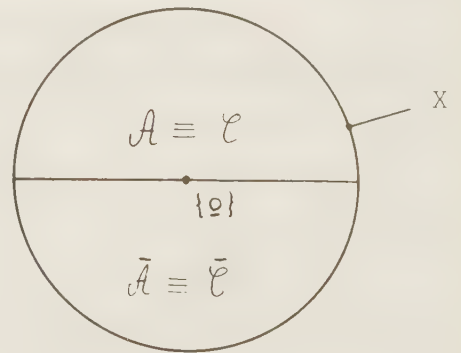
$$\mathcal{A} \oplus \bar{\mathcal{A}} = \mathcal{C} \oplus \bar{\mathcal{C}} = X \quad (\text{II.22})$$

et compte tenu de (II.20), on aura $\mathcal{A} \subseteq \mathcal{C}$ (figure 2). Dans le cas relativement fréquent où $\mathcal{A} \equiv \mathcal{C}$, la structure de l'espace X prend la forme simplifiée de la figure 3.



Cas discret
 Φ singulière

Fig. 2



Cas continu ou discret
 Φ régulière

Fig. 3

Les deux décompositions $\mathcal{C} \oplus \bar{\mathcal{C}}$ et $\mathcal{A} \oplus \bar{\mathcal{A}}$ concernant des problèmes distincts, il n'apparaît pas utile d'approfondir la structure globale

définie à la figure 2. Nous allons plutôt préciser séparément chacune d'entre-elles.

- Décomposition $\mathcal{A} \oplus \bar{\mathcal{A}}$

Reprenons la définition de l'accessibilité. Il est clair qu'un état n'est accessible que si, et seulement si, il appartient à l'espace engendré par les colonnes de $\Delta_{\bar{t}}^+$, (cf. aussi proposition II.4). Par suite, on énoncera le ...

Lemme II.1 : Le sous-espace accessible $\mathcal{A}$ s'identifie à $R(\Delta_{\bar{t}}^-)$.

Poursuivons en adoptant arbitrairement les notations relatives aux systèmes continus.

Par conséquent $\Delta_{\bar{t}}^- = \Delta_{\bar{t}}^-(A, B) = [B, AB, \dots, A^{\bar{t}-1}B]$. Considérons un état $\underline{x} \in \mathcal{A} \subseteq \mathbb{R}^n$, montrons alors que $A \underline{x} \in \mathcal{A}$. Par hypothèse :

$$\underline{x} \in R[B, AB, \dots, A^{\bar{t}-1}B] \quad (\text{II.23})$$

$$\Rightarrow A \underline{x} \in R[AB, A^2B, \dots, A^{\bar{t}}B] \quad (\text{II.24})$$

Or, d'après la définition de $\bar{t}$ (voir aussi proposition II.6), il en résulte que :

$$R[AB, A^2B, \dots, A^{\bar{t}}B] \subseteq R[B, AB, \dots, A^{\bar{t}-1}B] \quad (\text{II.25})$$

$$\Rightarrow A \underline{x} \in \mathcal{A} \quad (\text{II.26})$$

Formalisons :

Lemme II.2 : Le sous-espace accessible $\mathcal{A} \equiv R[\Delta_{\bar{t}}^-(A, B), (\equiv R[\Delta_{\bar{t}}^-(\Phi, \Gamma)])$ est invariant par rapport à A (Φ).

Considérons maintenant un système Σ qui ne soit pas complètement accessible, c'est-à-dire tel que :

$$\rho[\Delta_{\bar{t}}^-(A, B)] = n_1 < n \quad (\text{II.27})$$

Effectuons un changement de base dont les n_1 premières colonnes de la matrice de passage soient n_1 vecteurs (colonnes) linéairement indépendants, extraits de $\Delta_{\bar{t}}^-(A, B)$. Un résultat classique [7] (p.49) affirme alors que la matrice A exprimée dans cette nouvelle base, soit $\tilde{A}$, est de la forme suivante (en raison du lemme II.2) :

$$\tilde{A} = \begin{array}{c} \begin{array}{c} \xrightarrow{n_1} \\ \uparrow n_1 \end{array} \left[\begin{array}{cc|c} \tilde{A}_{a1} & \dots & \tilde{A}_{12} \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{A}_{\tilde{a}} \end{array} \right] \end{array} \quad (\text{II.28})$$

† On entend par là $\Delta_{\bar{t}}^-(\Phi, \Gamma)$ ou $\Delta_{\bar{t}}^-(A, B)$!

De plus, dans cette base tout $\underline{x} \in X$ se décompose d'une manière unique

$$\underline{x} \rightarrow \tilde{\underline{x}} = \tilde{\underline{x}}_a + \tilde{\underline{x}}_{\bar{a}} = \begin{bmatrix} \tilde{\underline{x}}_a \\ \vdots \\ 0 \end{bmatrix} \uparrow n_1 + \begin{bmatrix} 0 \\ \vdots \\ \tilde{\underline{x}}_{\bar{a}} \\ \vdots \\ 0 \end{bmatrix} \uparrow n-n_1 \quad (\text{II.29})$$

$$\text{soit } \tilde{\underline{x}}_a \in \tilde{A} \text{ et } \tilde{\underline{x}}_{\bar{a}} \in \tilde{A}.$$

Appliquons ce résultat aux colonnes de B, appartenant nécessairement à $\tilde{A}$, d'après le lemme II.1 ; il vient par suite :

$$\tilde{B} = \begin{bmatrix} \tilde{B}_a \\ \vdots \\ 0 \end{bmatrix} \uparrow n_1 \quad (\text{II.30})$$

On vient de redémontrer un résultat bien connu, à savoir :

Théorème II.3 : Etant donné un système Σ dont la paire (A,B) n'est pas complètement accessible, il existe un changement de base, tel que Σ se présente sous la forme :

$$\begin{bmatrix} \tilde{\underline{x}}_a \\ \vdots \\ \tilde{\underline{x}}_{\bar{a}} \end{bmatrix} = \begin{bmatrix} \tilde{A}_a & \tilde{A}_{12} \\ 0 & \tilde{A}_{\bar{a}} \end{bmatrix} \cdot \begin{bmatrix} \tilde{\underline{x}}_a \\ \vdots \\ \tilde{\underline{x}}_{\bar{a}} \end{bmatrix} + \begin{bmatrix} \tilde{B}_a \\ \vdots \\ 0 \end{bmatrix} \underline{u} \quad (\text{II.31})$$

et tel que le sous-système :

$$\dot{\tilde{\underline{x}}}_a = \tilde{A}_a \tilde{\underline{x}}_a + \tilde{B}_a \underline{u} \quad (\text{II.32})$$

soit accessible.

Remarque : Une version discrète de ce théorème s'énonce en changeant (A,B) en ($\hat{A}$, $\hat{B}$).

Adjoignons à (II.31) l'équation de sortie de Σ , notée ici :

$$\underline{y} = [\tilde{C}_a \mid \tilde{C}_{\bar{a}}] \begin{bmatrix} \tilde{\underline{x}}_a \\ \vdots \\ \tilde{\underline{x}}_{\bar{a}} \end{bmatrix} + D \underline{u} \quad (\text{II.33})$$

Il est clair d'après la structure de A et B (II.31) que l'état $\tilde{\underline{x}}_a$ évolue indépendamment de l'entrée. Par conséquent, la contribution de l'entrée sur la sortie, via l'état du système, ne se fera que par l'intermédiaire de $\tilde{\underline{x}}_{\bar{a}}$, d'où le ...

Théorème II.4 [1] Etant donné un système dynamique de dimension n non complètement accessible, il existe alors un sous-système de dimension $n_1 < n$ (II.31) - (II.32) accessible et ayant même relation d'entrée/sortie (matrice de transfert ou de réponses impulsionnelles).

Remarque : Là encore, on pourrait énoncer une version discrète en remplaçant (A,B,C) par (Φ, Γ, H) .

Ces deux théorèmes sont particulièrement importants parce qu'ils introduisent en termes formels le problème de réduction des réalisations. Des résultats analogues seront obtenus § 4, dans le contexte dual de l'observabilité, et ce sont ces deux développements qui conduiront à la propriété fondamentale d'irréductibilité des réalisations.

Pour l'instant, nous abordons la décomposition $\mathcal{C} \oplus \bar{\mathcal{C}}$. Nous le ferons brièvement en raison de l'aspect marginal que représentent, pour nous, ces résultats, puisqu'essentiellement orientés vers les problèmes de commande.

- Décomposition $\mathcal{C} \oplus \bar{\mathcal{C}}$

Remarquons d'abord qu'étant donné un système commandable ($\dim \mathcal{C} = n$), on peut se trouver dans le cas où $\rho[\Delta_{\mathcal{C}}] = n_1 < n$, (système discret et Φ singulière). Par suite, ce système, bien que commandable, admet néanmoins une décomposition (II.31) puisque non accessible et peut donc être réduit à (II.32) en tant que réalisation. Ceci justifierait, s'il en était encore besoin, que, des deux concepts introduits jusqu'ici, c'est l'accessibilité qui intervient dans les problèmes de réalisation.

Supposons maintenant que Σ ne soit pas complètement commandable ($\dim \mathcal{C} < n$). Définissons rapidement un problème de commande en indiquant qu'il s'agit de placer les pôles du système bouclé d'une manière donnée. On sait qu'on ne peut agir (nous dirions presque par définition) que sur les pôles relatifs aux états commandables. L'instabilité possible du système résultant a précisément motivé l'introduction du qualificatif *stabilisable*.

Définition II.3 [8] On dit que la paire (A,B) est *stabilisable*, s'il existe une matrice K telle que le système bouclé $\dot{x} = (A+BC)x$ soit asymptotiquement stable.

Remarque : Il est clair que :

$$\begin{aligned} \Sigma \text{ commandable} & \Rightarrow \Sigma \text{ stabilisable} \\ \Sigma \text{ asymptotiquement stable} & \Rightarrow \Sigma \text{ stabilisable} \end{aligned}$$

Nous terminerons en rappelant deux critères de stabilisabilité.

Théorème II.5 : Soit $\mathcal{J} \subseteq X$ le sous-espace des états instables, alors Σ est stabilisable si, et seulement si, $\mathcal{J} \subseteq \mathcal{C}$.

Remarque : Ceci signifie, plus simplement, que les états instables sont commandables.

Un cas particulier important est celui où $\mathcal{C} \equiv \mathcal{A}$. Compte tenu du théorème II.3, on a alors le ...

Corollaire II.8 : Lorsque les états commandables sont aussi accessibles, le système Σ est stabilisable si, et seulement si, le sous-système $\dot{\underline{x}}_{\bar{a}} = A_{\bar{a}}^{\nu} \underline{x}_{\bar{a}}$ est asymptotiquement stable.

Remarque : Là encore tous ces résultats se transposent aux systèmes discrets en remplaçant (A,B) par (Φ, Γ).

4 - OBSERVABILITE ET RECONSTRUCTIBILITE

L'observabilité et la restructibilité sont au couple état-sortie ce que sont l'accessibilité et la commandabilité au couple entrée-état. En termes plus précis, le problème d'*observation* consiste à déterminer l'évènement *présent* ($\underline{x}, \tau$) à partir de la connaissance de la sortie *future*. Inversement, la *reconstruction* de l'état consiste à déterminer l'évènement *présent*, par la connaissance de la sortie *passée* [2].

Cette symétrie dans le temps, que l'on avait déjà soulignée lors de la définition de la commandabilité et de l'accessibilité, se retrouve donc intégralement ici. Il ressort, en fait, de ces quatre concepts, pris deux à deux, une certaine analogie qu'il serait d'ailleurs plus exact de qualifier de complémentarité. En effet, rappelons que l'existence d'un régulateur est liée à la commandabilité du système, mais sa mise en oeuvre suppose que l'on puisse déterminer l'état à partir de mesures (nécessairement) passées, en l'occurrence sur la sortie, c'est-à-dire reconstruire l'état.

L'association accessibilité - observabilité se référera, quant-à elle, au problème de réalisation minimale. Mais n'anticipons pas sur ce point que nous reprendrons en détail au § 5. Pour l'instant, formalisons ces deux nouveaux concepts par les ...

Définition II.4 : Un évènement ($\underline{x}, \tau$) est *observable* si, et seulement si, il existe un instant $t \geq \tau$, fini, tel que cet évènement puisse être déterminé d'une façon unique à partir de la sortie $\underline{y}[\tau, t]$ du système, en régime libre ($\underline{u}[\tau, t] \equiv \underline{0}$).

... et ...

Définition II.5 : Un évènement ($\underline{x}, \tau$) est *reconstructible* si, et seulement si, il existe un instant $s \leq \tau$, fini, tel que cet évènement puisse être déterminé d'une façon unique à partir de la sortie $\underline{y}[s, \tau]$ du système, en régime libre ($\underline{u}[s, \tau] \equiv \underline{0}$).

La condition régime libre, introduite ici, permet de mettre en évidence, clairement, le fait que l'observabilité et la restructibilité ne sont des propriétés du système, relatives qu'au couple état-sortie. Cet énoncé n'est néanmoins nullement restrictif, puisque dans le cas général (régime forcé), il suffit de distinguer, dans la sortie, la contribution de l'entrée de celle de l'état initial, pour se ramener au problème précédent.

On a en effet :

$$\underline{y}(t_2) - C \int_{t_1}^{t_2} e^{A(t_2-\tau)} B \underline{u}(\tau) d\tau - D \underline{u}(t_2) = C e^{A(t_2-t_1)} \underline{x}(t_1) \quad (II.34)$$

ou dans le cas discret :

$$\underline{y}_{t_2} - H \sum_{\tau=0}^{t_2-1} \phi^{t_2-\tau-1} \Gamma \underline{u}_{\tau} - D \underline{u}_{t_2} = H \phi^{t_2-t_1} \underline{x}_{t_1} \quad (II.35)$$

expressions que l'on écrit encore :

$$\underline{y}^*(t_2) = C e^{A(t_2-t_1)} \underline{x}(t_1) \quad \text{et} \quad \underline{y}_{t_2}^* = H \phi^{t_2-t_1} \underline{x}_{t_1} \quad (II.36)$$

Un système Σ observable (restructible) en régime libre, l'est donc en régime forcé et réciproquement.

Dans ce qui suit, on va mettre en évidence puis raisonner sur le parallèle (*dualité*) qui existe entre les résultats relatifs à l'accessibilité et la commandabilité d'une part, et, l'observabilité et la restructibilité d'autre part.

Le cas continu étant déjà très largement traité [1], [2], [3], nous avons choisi d'aborder le problème essentiellement sous l'angle discret, qui présente, nous l'avons déjà vu § 3, certains aspects particuliers.

Ces considérations expliquent ainsi le plan adopté dans les deux paragraphes suivants où les développements relatifs au cas où $T = R$ ne sont effectués que d'une façon allusive.

4.1 - Critère d'observabilité

On notera sans restreindre le problème, $t_1 = \tau = 0$ et $t_2 = t$, de sorte qu'il s'agit de préciser dans quelles conditions on peut déterminer $\underline{x}_0$ d'une façon unique, connaissant $\underline{y}[0, t]$. De l'équation (II.36), il vient :

$$\underline{y}(t) = C e^{At} \underline{x}(0) \quad \text{et} \quad \underline{y}_t = H \phi^t \underline{x}_0 \quad (II.37)$$

Cette dernière relation définit une application linéaire de X dans Y , que l'on notera :

$$X \xrightarrow{h} Y \quad (II.38)$$

L'unicité caractéristique de l'état initial implique et réciproquement, que le noyau de h se réduise au vecteur nul, c'est-à-dire soit de dimension nulle. Nous formulerons ce résultat par le ...

Théorème II.6 : Un système dynamique Σ est observable si, et seulement si, l'application h (II.38) est injective.

Explicitons dans le cas discret ; on déduit de (II.37) :

$$t \times p \begin{matrix} \uparrow \\ \left[\begin{array}{c} y_0 \\ \vdots \\ y_{t-1} \end{array} \right] \\ \downarrow \end{matrix} = \begin{bmatrix} H \\ \vdots \\ H \Phi^{t-1} \end{bmatrix} \underline{x}_0 \quad (\text{II.39})$$

que l'on écrit encore :

$$\underline{y} = \Lambda_t(\Phi, H) \underline{x}_0 \quad (\text{II.40})$$

avec par définition :

$$\Lambda_t = \Lambda_t(\Phi, H) = \begin{bmatrix} H \\ \vdots \\ H \Phi^{t-1} \end{bmatrix}$$

En se référant aux propositions (II.5) à (II.7) et en notant $\bar{t} = \min(n - \rho(H) + 1, \bar{n})$ on en déduit immédiatement le ...

Corollaire II.9 : Un système Σ est observable si, et seulement si, $\rho(\Lambda_{\bar{t}}) = n$.

Indiquons brièvement pour justifier la terminologie générale de ce corollaire, que dans le cas continu on est conduit à une relation analogue à (II.40) en dérivant $(t-1)$ fois $\underline{y}(0)$. La condition h injective est alors encore caractérisée par $\rho[\Lambda_{\bar{t}}(A, C)] = n$.

Rapprochons ce résultat des corollaires (II.2) et (II.3), on constate que :

$$\Delta_{\bar{t}}(., *) = \Lambda_{\bar{t}}^T(., *^T) \quad (\text{II.41})$$

C'est cette correspondance que l'on nomme *dualité*, terme dont on élargit la signification à travers la ...

Définition II.6 : Le *dual* du système $\Sigma(\Phi, \Gamma, H, D)$, (ou $\Sigma(A, B, C, D)$) est le système $\Sigma^*(\Phi^T, H^T, \Gamma^T, D^T)$, respectivement $\Sigma^*(A^T, C^T, B^T, D^T)$.

Dans ces conditions, il est clair que :

$$\Delta_{\bar{t}}^{*T} = \Lambda_{\bar{t}} \quad \text{et} \quad \Lambda_{\bar{t}}^{*T} = \Delta_{\bar{t}} \quad (\text{II.42})$$

d'où le premier théorème de dualité.

Théorème II.7 :

Le système Σ est accessible si, et seulement si, Σ^* est observable.

Le système Σ est observable si, et seulement si, Σ^* est accessible.

4.2 - Critère de restructibilité

L'étude de la restructibilité peut donc se déduire, d'après (II.42), par dualité des résultats acquis dans le domaine de la commandabilité ; ce qui nous conduit au deuxième théorème de dualité :

Théorème II.8 :

Le système Σ est restructible si, et seulement si, Σ^* est commandable.

Le système Σ est commandable si, et seulement si, Σ^* est restructible.

Remarque : Par dualité, on a bien sûr :

$$\begin{aligned} \Sigma \text{ observable} & \Rightarrow \Sigma \text{ restructible} \\ \Sigma \text{ restructible} & \not\Rightarrow \Sigma \text{ observable} \end{aligned} \quad (\text{II.43})$$

Ce résultat de synthèse étant acquis, on propose néanmoins une approche directe du cas discret, en raison de l'aspect constructif de la démonstration.

En conservant les notations précédentes, le problème est donc de déterminer d'une façon unique l'état $\underline{x}_t$, connaissant $\underline{y}[0, t]$. Si Φ est régulière, le problème est trivial, puisqu'il y a bijection entre $\underline{x}_0$ et $\underline{x}_t$ ($\underline{x}_t = \Phi^t \underline{x}_0$).

Supposons donc Φ singulière. La difficulté, dans ce cas, provient, d'une part, qu'on ne peut écrire une relation explicite entre $\underline{x}_t$ et $\underline{y}_0, \underline{y}_1, \dots, \underline{y}_t$, du type (II.39), et d'autre part, qu'il peut exister plusieurs $\underline{x}_0$ conduisant au même $\underline{x}_t$. Ceci nous amène à énoncer le problème de restructibilité de la manière suivante (équivalente) :

"L'état $\underline{x}_t$ sera restructible si, et seulement si, il existe au moins un $\underline{x}_0$ vérifiant (II.39), et tel qu'il lui corresponde un et un seul $\underline{x}_t (= \Phi^t \underline{x}_0)$ ".

Ceci revient à définir une injection h' de $X' = R(\Phi^t)$ dans Y , d'où le dual du théorème II.2 :

Théorème II.8 : Un système Σ est restructible si, et seulement si, l'application $X' \xrightarrow{h'} Y$ est injective.

Revenons au cas discret avec l'équation (II.38) :

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{t-1} \end{bmatrix} \underline{x}_0$$

L'existence d'un $\underline{x}_0$ est assurée par la définition même des $y_0, y_1, \dots, y_{t-1}$. Soit $\underline{x}'_0$ une autre solution, c'est-à-dire $\underline{y} = \Lambda_t \underline{x}'_0$, d'après (II.40). En posant $\underline{x}'_0 = \underline{x}_0 + \underline{\delta x}$, ceci implique que $\Lambda_t \underline{\delta x} = \underline{0}$. Puisque $\underline{\delta x} \neq \underline{0}$ a priori, on aura donc $\underline{\delta x} \in N(\Lambda_t)$. Il reste à déterminer la condition pour que $\underline{x}_t$ soit unique, c'est-à-dire pour que $\Phi^t \underline{\delta x} = \underline{0}$ ou encore $\underline{\delta x} \in N(\Phi^t)$. Il est clair que :

$$(\Lambda_t \underline{\delta x} = \underline{0} \Rightarrow \Phi^t \underline{\delta x} = \underline{0}) \iff N(\Lambda_t) \subseteq N(\Phi^t) \quad (\text{II.44})$$

De cette dernière équation, il vient :

$$R(\Lambda_t^T) \supseteq R(\Phi^{tT}) \quad (\text{II.45})$$

qui est effectivement la forme duale de :

$$R(\Phi^t) \subseteq R(\Lambda_t)$$

d'où finalement le dual du corollaire II.7 :

Corollaire II.10 : La paire (Φ, H) est reconstructible si, et seulement si, pour un $t \leq \bar{t}$ (II.45) est vérifié.

Exemple et contre exemple

Soit le système dynamique Σ :

$$\underline{x}_{t+1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \underline{x}_t$$

$$\underline{y}_t = (1 \quad 0) \underline{x}_t$$

Donnons nous arbitrairement $\underline{x}_0 = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, on en déduit successivement :
 $y_0 = \alpha$; $\underline{x}_1 = \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$ et $y_1 = 0$

D'après le corollaire II.9, la paire (Φ, H) n'est pas observable puisque :

$$\rho \begin{bmatrix} H \\ H\Phi \end{bmatrix} = \rho \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 1 < 2$$

En effet, étant donné y_0, y_1 , il existe une infinité de solutions $\underline{x}_0$, vérifiant (II.39).

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} H \\ H\Phi \end{bmatrix} \underline{x}_0 \Leftrightarrow \begin{bmatrix} \alpha \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \lambda \end{bmatrix}, \quad \forall \lambda$$

Par ailleurs, si l'on se réfère au théorème 1 de l'article de Dorato [9], on en concluerait que la paire (Φ, H) n'est pas non plus reconstructible, ce qui est inexact. Reprenons le $\underline{x}_0$ trouvé $\begin{bmatrix} \alpha \\ \lambda \end{bmatrix}$ il lui correspond $\underline{x}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$ qui est unique car indépendant de λ ; donc la paire (Φ, H) est reconstructible, ce que confirme le corollaire II.10 :

$$\Lambda_1^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; R(\Lambda_1^T) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ et } \Phi^T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow R(\Phi^T) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Remarque : Le théorème de Dorato n'exprime, en fait, qu'une condition suffisante, à moins que Φ soit régulière.

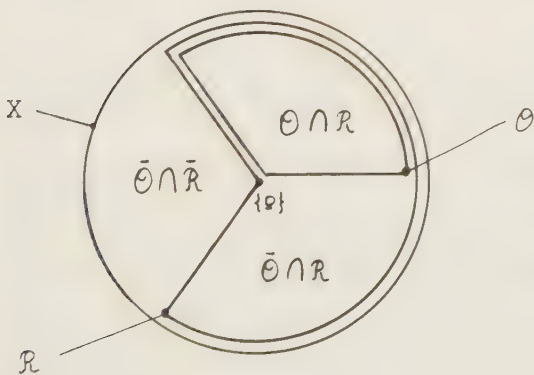
4.3 - Décomposition de l'espace d'état 2ème partie

Adoptons les notations suivantes :

- $\mathcal{O}$ ensemble des états observables - dual de $\mathcal{A}$
- $\bar{\mathcal{O}}$ ensemble des états non observables - dual de $\bar{\mathcal{A}}$
- $\mathcal{R}$ ensemble des états reconstructibles - dual de $\mathcal{C}$
- $\bar{\mathcal{R}}$ ensemble des états non reconstructibles - dual de $\bar{\mathcal{C}}$

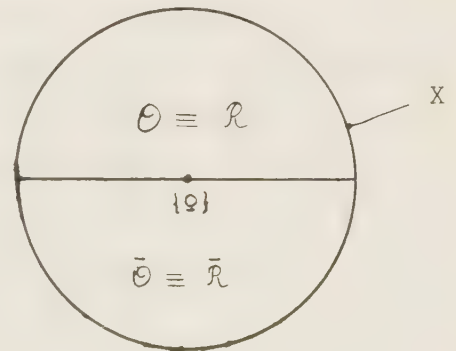
et remarquons d'après (II.43) que $\mathcal{O} \subseteq \mathcal{R}$ inégalité duale de $\mathcal{A} \subseteq \mathcal{C}$.

Toujours par dualité, on obtient à partir des figures 2 et 3, la structure de X relative aux deux nouvelles décompositions $\mathcal{O} \oplus \bar{\mathcal{O}}$ et $\mathcal{R} \oplus \bar{\mathcal{R}}$, figures 4 et 5.



Cas discret
 Φ singulière

Fig. 4



Cas continu ou discret
 Φ régulière

Fig. 5

- Décomposition $\theta \oplus \bar{\theta}$

Nous savons, lemme II.1, que $\mathcal{A} \equiv R(\Lambda_{\bar{t}}^T)$, en prenant le dual des deux membres, c'est-à-dire θ et $R[\Lambda_{\bar{t}}^T]$, il en découle le ...

Lemme II.3 : Le sous-espace observable θ s'identifie à $R(\Lambda_{\bar{t}}^T)$.

Appliquons maintenant le lemme II.2 au système dual, ceci donne (cf. II.25) :

$$R[A^T C^T, A^{T^2} C^T, \dots, A^{T^{\bar{t}}} C^T] \subset R[C^T, A^T C^T, \dots, A^{T^{\bar{t}-1}} C^T]$$

$$\Leftrightarrow N \begin{bmatrix} C & A \\ C & A^2 \\ \vdots & \vdots \\ C & A^{\bar{t}} \end{bmatrix} \supset N \begin{bmatrix} C \\ C & A \\ \vdots & \vdots \\ C & A^{\bar{t}-1} \end{bmatrix} \quad (\text{II.46})$$

Cette dernière expression signifie que si $\underline{x} \in \bar{\theta}$ alors $A \underline{x} \in \bar{\theta}$; en d'autres termes :

Lemme II.4 : Le sous-espace *non-observable* $\bar{\theta} \equiv N[\Lambda_{\bar{t}}^T(A, C)]$, (ou $\equiv N[\Lambda_{\bar{t}}^T(\Phi, H)]$) est invariant par rapport à A (ou à Φ).

Considérons maintenant un système Σ non observable, c'est-à-dire tel que Σ^* ne soit pas accessible. Appliquons à ce dernier les théorèmes II.3 et II.4, il vient :

$$\left\{ \begin{aligned} \begin{bmatrix} \dot{\tilde{x}}_a \\ \vdots \\ \tilde{x}_a \end{bmatrix}^* &= \begin{bmatrix} \tilde{A}_a^T & \tilde{A}_{12}^T \\ 0 & \tilde{A}_a \end{bmatrix}^* \begin{bmatrix} \tilde{x}_a \\ \vdots \\ \tilde{x}_a \end{bmatrix}^* + \begin{bmatrix} \tilde{C}_a^T \\ \vdots \\ 0 \end{bmatrix}^* \underline{u}^* \\ \underline{y}^* &= [\tilde{B}_a^T \quad \tilde{B}_{\bar{a}}^T]^* \begin{bmatrix} \tilde{x}_a \\ \vdots \\ \tilde{x}_a \end{bmatrix}^* + D^T \underline{u}^* \end{aligned} \right. \quad (\text{II.47})$$

Transposons au système original :

$$\left\{ \begin{aligned} \begin{bmatrix} \dot{\tilde{x}}_o \\ \vdots \\ \tilde{x}_o \end{bmatrix} &= \begin{bmatrix} \tilde{A}_o & 0 \\ \tilde{A}_{12} & \tilde{A}_o \end{bmatrix} \begin{bmatrix} \tilde{x}_o \\ \vdots \\ \tilde{x}_o \end{bmatrix} + \begin{bmatrix} \tilde{B}_o \\ \vdots \\ \tilde{B}_o \end{bmatrix} \underline{u} \\ \underline{y} &= [\tilde{C}_o \quad 0] \begin{bmatrix} \tilde{x}_o \\ \vdots \\ \tilde{x}_o \end{bmatrix} + D \underline{u} \end{aligned} \right. \quad (\text{II.48})$$

En conséquence :

Théorème II.9 : Etant donné un système Σ , non complètement observable, il existe un changement de base, tel qu'il soit de la forme (II.48) et tel que le sous-système $(\tilde{A}_o, \tilde{B}_o, \tilde{C}_o, D)$ soit observable et ait même relation d'entrée sortie (matrice de transfert).

Remarque : Les notations "continues" ne sont évidemment pas restrictives.

- Décomposition $\mathcal{R} \oplus \bar{\mathcal{R}}$

En raisonnant soit directement sur le résultat de ce paragraphe, soit sur ceux du § 3.3, par dualité, on constate qu'un système restructurable peut admettre une décomposition (II.48). Par conséquent, accessibilité et observabilité sont des propriétés qui concernent les problèmes de réalisations (minimales), alors que commandabilité et restructurabilité sont des caractéristiques relevant de la commande et de la régulation des systèmes. Dans ce contexte, si un système n'est pas complètement restructurable, les états correspondants peuvent introduire une erreur croissante au cours du temps. Lorsqu'il n'en n'est rien, on dit que le système est *déTECTABLE*.

Définition II.7 : Un système Σ est détectable si, et seulement si, son dual est stabilisable. On en déduit immédiatement d'après le théorème II.5 le

Corollaire II.12 : Un système est détectable si, et seulement si,

$$\mathcal{J} \subset \mathcal{R}$$

Remarque : Ceci signifie que Σ est détectable lorsque les états instables sont restructurables.

Lorsque l'on a $\Theta \equiv \mathcal{R}$ la décomposition II.48 fournit un critère explicite que nous formalisons dans le ...

Corollaire II.13 : Le système dynamique Σ est détectable si, et seulement si, le système $\tilde{\mathbf{x}}_0 = \tilde{\mathbf{A}}_0 \tilde{\mathbf{x}}_0$ est asymptotiquement stable.

Remarque : Ceci est toujours exact dans le cas discret, bien évidemment.

5 - STRUCTURE DE L'ESPACE D'ETAT DEFINI PAR L'ACCESSIBILITE ET L'OBSERVABILITE

On vient de constater (séparément) que seule la partie accessible ou observable d'un système intervenait effectivement dans sa relation d'entrée/sortie. De la combinaison de ces deux décompositions ($\mathcal{A} \oplus \bar{\mathcal{A}}$ et $\mathcal{O} \oplus \bar{\mathcal{O}}$) se déduit un des résultats fondamentaux du problème de réalisation, généralement connu sous le nom de "théorème de décomposition canonique", énonçant que seules les parties accessibles et observables contribuaient à définir la relation d'entrée/sortie du système.

Nous allons redémontrer rapidement ce théorème en adoptant arbitrairement les notations relatives aux systèmes continus. Introduisons pour cela les sous-espaces suivants :

- $X_1 = \mathcal{A} \cap \bar{\Theta}$ l'ensemble des états accessibles et non observables
 $X_2 = \mathcal{A} \cap \Theta$ l'ensemble des états accessibles et observables
 $X_3 = \bar{\mathcal{A}} \cap \bar{\Theta}$ l'ensemble des états ni accessibles, ni observables
 $X_4 = \bar{\mathcal{A}} \cap \Theta$ l'ensemble des états non accessibles, mais observables

On vérifie facilement que :

$$X_1 \oplus X_2 \oplus X_3 \oplus X_4 = X \quad (\text{II.49})$$

Notons $\Sigma(A, B, C, D)$ le système dynamique Σ exprimé dans la base naturelle de X . Effectuons un changement de base, en prenant comme nouvelle base de X , l'union des bases de X_i , $i = 1, \dots, 4$, et partitionnons les matrices A, B, C exprimées dans cette nouvelle base (soient $\tilde{A}, \tilde{B}, \tilde{C}$) suivant la dimension des X_i , $i = 1, \dots, 4$; c'est-à-dire :

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} & \tilde{A}_{14} \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} & \tilde{A}_{24} \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} & \tilde{A}_{34} \\ \tilde{A}_{41} & \tilde{A}_{42} & \tilde{A}_{43} & \tilde{A}_{44} \end{bmatrix} ; \quad \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \tilde{B}_3 \\ \tilde{B}_4 \end{bmatrix} \quad \text{et} \quad \tilde{C} = [\tilde{C}_1 \quad \tilde{C}_2 \quad \tilde{C}_3 \quad \tilde{C}_4] \quad (\text{II.50})$$

On sait que $\mathcal{A} = X_1 \oplus X_2$ (voir définitions) est invariant par rapport à A , de sorte que d'après (II.31) on aura : $\tilde{A}_{31} = \tilde{A}_{32} = \tilde{A}_{41} = \tilde{A}_{42} = 0$ et $\tilde{B}_3 = \tilde{B}_4 = 0$

Par ailleurs, selon le lemme (II.4), $\bar{\Theta} = X_1 \oplus X_3$ est également invariant par rapport à A , et en se référant à (II.48), on en déduit que : $\tilde{A}_{21} = \tilde{A}_{41} = \tilde{A}_{23} = \tilde{A}_{43} = 0$ et $\tilde{C}_1 = \tilde{C}_3 = 0$.

En conséquence Σ s'exprime dans la nouvelle base suivant la forme (II.51) dite canonique :

$$\left\{ \begin{array}{l} \begin{bmatrix} \tilde{x}_{a\bar{o}} \\ \tilde{x}_{ao} \\ \tilde{x}_{\bar{a}\bar{o}} \\ \tilde{x}_{\bar{a}o} \end{bmatrix} = \begin{bmatrix} \tilde{A}_{a\bar{o}} & \tilde{A}_{12} & \tilde{A}_{13} & \tilde{A}_{14} \\ 0 & \tilde{A}_{ao} & 0 & \tilde{A}_{24} \\ 0 & 0 & \tilde{A}_{\bar{a}\bar{o}} & \tilde{A}_{34} \\ 0 & 0 & 0 & \tilde{A}_{\bar{a}o} \end{bmatrix} \cdot \begin{bmatrix} \tilde{x}_{a\bar{o}} \\ \tilde{x}_{ao} \\ \tilde{x}_{\bar{a}\bar{o}} \\ \tilde{x}_{\bar{a}o} \end{bmatrix} + \begin{bmatrix} B_{a\bar{o}} \\ B_{ao} \\ 0 \\ 0 \end{bmatrix} \underline{u} \\ \\ \underline{y} = \begin{bmatrix} 0 & \tilde{C}_{ao} & 0 & \tilde{C}_{\bar{a}o} \end{bmatrix} \begin{bmatrix} \tilde{x}_{a\bar{o}} \\ \tilde{x}_{ao} \\ \tilde{x}_{\bar{a}\bar{o}} \\ \tilde{x}_{\bar{a}o} \end{bmatrix} + D \underline{u} \end{array} \right. \quad (\text{II.51})$$

En appliquant à (II.51) successivement le théorème II.4 et II.9, il en résulte que :

$$H(s) = C(sI - A)^{-1} B + D = \tilde{C}_{ao}(sI - \tilde{A}_{ao})^{-1} B_{ao} + D \quad (II.52)$$

Résultats que l'on résume dans le ...

Théorème II.10 : Etant donné un système dynamique Z , il existe toujours un changement de base tel que Z puisse se mettre dans la forme canonique (II.51). La relation d'entrée/sortie de Z est alors identique à celle du sous-système accessible et observable.

Remarque : Une version discrète du théorème s'obtient en remplaçant (A, B, C) par (Φ, Γ, H) .

6 - REALISATION MINIMALE D'UN SYSTEME DYNAMIQUE

6-1 L'approche classique

On considère d'abord le problème de réalisation au sens initial de la définition I.2 du I.3, c'est-à-dire la réalisation d'une représentation externe, (matrice de transfert par exemple).

La minimalité relève alors de la question qui consiste à savoir si l'on peut réduire la dimension de X sans changer la matrice de transfert de Z . Formalisons :

Définition II.8 : Un système dynamique de dimension n , Z_n est *réductible* si, et seulement si, il existe un système dynamique Z_{n_1} , $n_1 < n$ et ayant même matrice et transfert.

Revenons maintenant au théorème II.10 dont l'interprétation graphique, figure 6, illustre clairement le fait que :

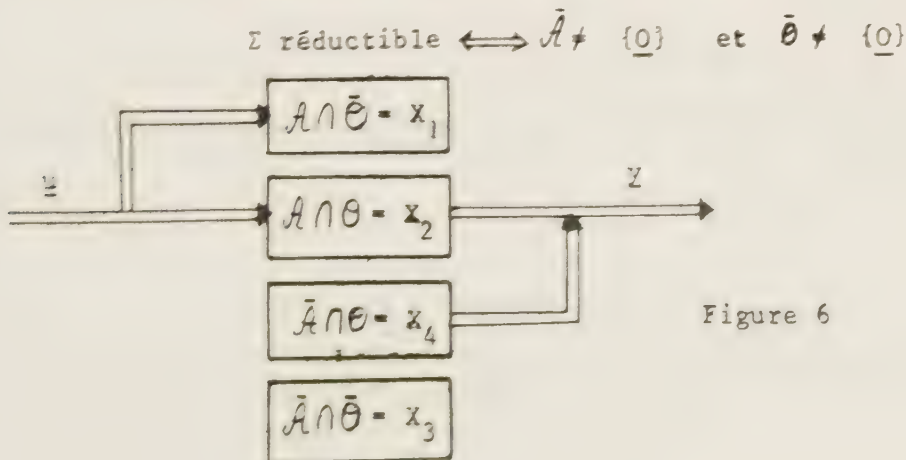


Figure 6

En d'autres termes :

Corollaire II.14 = Un système dynamique Σ est *irréductible* et donc *minimal*[†] si, et seulement si, il est accessible et observable.

6.2 - Réalisation d'entrées/sorties

Si l'on généralise la notion de réalisation comme nous l'avons fait au chapitre I (Définition I.8), il convient de faire quelques remarques. Les données du problème étant maintenant directement u et y , il est clair, d'après la Figure 6, que l'on pourra réaliser non seulement X_2 mais aussi X_4 . La réalisation minimale que l'on pourra construire, c'est-à-dire celle de dimension la plus petite compatible avec les données, pourra donc contenir des états inaccessibles.

Ici donc, minimal n'implique pas irréductible, à moins que l'état initial ne soit nul, tout au moins la partie inaccessible, ou qu'a priori, le système à partir duquel sont acquises les données soit lui-même accessible.

Avant de conclure ce chapitre, on voudrait exposer, dans son principe, une approche parallèle permettant de conduire au corollaire (II.14), sans avoir à développer des résultats intermédiaires tels que ceux exprimés par les théorèmes II.3, II.4, II.9 et II.10, dont la formulation fait nécessairement intervenir T.

6.3 - Un formalisme plus général

Reprenons la définition de base d'une réalisation. Soit un système linéaire invariant défini par son application f (ici linéaire) de U dans Y , et un système dynamique Σ caractérisé par deux applications g et h telles que :

$$U \xrightarrow{g} X \xrightarrow{h} Y \quad (II.53)$$

Définition II.8 : Le système Σ est une réalisation de f si, et seulement si, $f_{\Sigma} = f$, c'est-à-dire :

$$f = h \circ g \quad (II.54)$$

Poursuivons, en introduisant quelques concepts nouveaux, appropriés à ce développement.

[†] Minimal au sens de la définition I.5

Définition II.9 : [2] Soit deux ensembles U et Y , et une application f de U dans Y . On dit que f est *factorisé* par rapport à l'ensemble X si, et seulement si, il existe deux applications g et h respectivement de U dans X et de X dans Y , telle que $f = h \circ g$.

Construire une réalisation d'un système, décrit par sa relation d'entrée/sortie, consiste donc à factoriser cette dernière par rapport à un espace d'état.

Définition II.10 : [2] Une factorisation est *canonique* si, et seulement si, g est surjectif et h injectif.

Si on fait le parallèle avec les résultats des § 3 et 4, canonique est donc à rapprocher d'irréductible, puisque g surjectif signifie accessible (théorème II.1), et h injectif observable (théorème II.6). Il ne reste plus qu'à démontrer que la factorisation canonique est celle pour laquelle la dimension de X est minimale, ou irréductible.

Lemme II.5 $X = R(f)^{\dagger}$

En effet soit $f = h \circ g \Rightarrow h(X) \supseteq R(f)$ sinon on aurait $f \neq h \circ g$. En prenant $X = R(f)$ il vient $h(X) \subseteq X$. Donc finalement $h(X) = X$ ce qui caractérise une injection (Noyau de dimension nulle). Par ailleurs, on a $g(U) \subseteq X$. Si $g(U) \subsetneq X$ on aura $h \circ g(U) \subsetneq h(X) = R(f) \Rightarrow f \neq h \circ g$; donc $g(U) = X$, c'est-à-dire g surjectif.

Corollaire II.15 $\dim X \text{ canonique} = \dim R(f)$.

Théorème II.11 La dimension de X est minimale (irréductible) si, et seulement si, la factorisation de f , par rapport à X est canonique.

Démonstration Soit $f = h \circ g$ une factorisation canonique, et $f = h' \circ g'$ une factorisation quelconque. Puisque $f = h' \circ g'$ on a $h'(X') \supseteq R(f) \Rightarrow$

$$\dim X' \geq \dim h'(X') \geq \dim R(f) = \dim X \quad (\text{II.55})$$

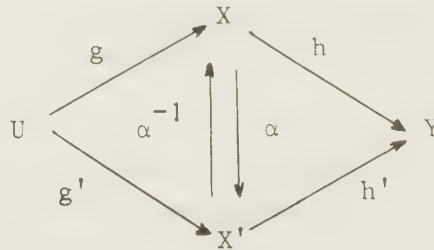
donc canonique $\Rightarrow$ minimale.

$\dagger R(f)$ désigne par extension de $R(M)$, (M matrice), l'ensemble $f(U)$ engendré par f .

Supposons maintenant $\dim X' = \dim X \implies \dim X' = \dim h'(X') = \dim R(f)$. Ceci signifie que le noyau de h est de dimension nulle, c'est-à-dire h' injectif. Enfin si $g'(U) \subset X'$, on aura, là encore, $h \circ g(U) \subset h'(X') = R(f) \implies f \neq h \circ g$; d'où nécessairement $g'(U) = X'$, c'est-à-dire g' surjectif. Conclusion : minimale $\implies$ canonique.

Remarque : Une application f est donc de dimension finie si, et seulement si, elle est factorisable par rapport à un espace de dimension fini (*Réalisabilité*, cf théorème I.1).

Dans ce contexte, l'unicité, modulo un changement de base, des réalisations irréductibles se démontrent rapidement en faisant appel au Lemme de Zeiger [10] sur les diagrammes commutatifs. Il en résulte, qu'étant donné deux factorisations *canoniques* $f = h \circ g = h' \circ g'$, il existe une bijection α *unique* telle que le diagramme ci-après commute.



ce qui implique :

$$g' = \alpha \circ g \quad \text{et} \quad h = h' \circ \alpha \iff h' = h \circ \alpha^{-1}$$

résultat que l'on rapproche immédiatement de (I.27), en notant P la matrice régulière associée à α , et qui démontre bien sûr (I.28).

D'autre part, il est maintenant clair que le théorème I.2 ne peut avoir de réciproque. En effet, le fait que α soit bijectif implique que X et X' aient le même nombre d'éléments ce qui ne sera pas a priori vérifié si les applications g et h ne sont pas respectivement surjectives et injectives. En termes plus classiques, intégrons successivement une réalisation irréductible de dimension n comme sous système d'une réalisation seulement observable de dimension $n' > n$ puis d'une réalisation seulement accessible de même dimension n' il est clair qu'il n'existe pas de matrice P permettant de passer de l'une à l'autre de ces deux réalisations.

Signalons, pour terminer, que cette seconde approche, un peu plus abstraite, fait en réalité appel à une terminologie très significative. Cette démarche, qui repose sur la *factorisation* de f , est exactement celle qu'empruntent les algorithmes classiques opérant sur les paramètres de Markov. Rappelons en effet que f est entièrement caractérisé par la suite infinie de

ces paramètres généralement disposés de façon à former une matrice de Hankel[†] de dimension infinie.

Soit $\mathcal{F}$ cette matrice explicitée ici en supposant pour simplifier les notations que $m = p = 1$.

$$\mathcal{F}_{\infty \times \infty} = \begin{bmatrix} f_1 & f_2 & \dots & f_t & & \\ f_2 & f_3 & & & & \\ \vdots & \vdots & & & \dots & \\ \vdots & \vdots & & & & \\ f_t & f_{t+1} & \dots & f_{2t-1} & & \\ & \vdots & & & & \\ & \vdots & & & & \end{bmatrix} \quad (\text{II.56})$$

Effectuer une factorisation canonique de f est donc équivalente à la définition de deux matrices $M_{\infty \times n}$ et $N_{n \times \infty}$ de rang n telles que $\mathcal{F} = M.N$ (n n'est autre que le rang de f , inégalité de Sylvester). On fera le parallèle $\rho(M) = n \iff h$ surjectif, $\rho(N) = n \iff g$ surjectif ; $n = \rho(\mathcal{F}) \iff n$ (canonique) = $\dim(f)$. En se référant aux définitions de Δ et Λ (II.13) et (II.40), on vérifie facilement que d'une part,

$$\mathcal{F} = \Lambda_{\infty} \Delta_{\infty} \quad (\text{II.57})$$

et d'autre part, que canonique est synonyme d'accessible et observable, c'est-à-dire une fois encore irréductible.

7 - CONCLUSION

De cette étude de la structure et des propriétés de l'espace d'état X , il convient de retenir les points essentiels suivants :

Une réalisation est irréductible si, et seulement si, le système dynamique Σ correspondant est accessible et observable. En d'autres termes, construire une réalisation irréductible consiste à factoriser une application f de U dans Y par rapport à un espace d'état X de dimension minimale. Ce résultat est obtenu en décomposant f sous la forme $h \circ g$ où h et g sont deux applications respectivement de X dans Y , injective et de U dans X , surjective. On retiendra également que l'ensemble des réalisations irréductibles de f forme une classe d'équivalence, modulo un changement de base. Rappelons enfin la nuance à faire

[†] Une matrice de Hankel est par définition une matrice d'éléments $[*]_{ij} = \bullet_{i+j}$
c.a.d. $[*]_{ij} = C_{i+j}^{te}$ si $i+j = C_{i+j}^{te}$

entre minimalité et irréductibilité lorsqu'il s'agit de déterminer une réalisation directement à partir d'entrées/sorties quelconques.

Dans un but de synthèse, les principaux résultats développés au cours de ce chapitre sont regroupés dans le tableau ci-après.

Σ système dynamique invariant (modèle d'état)		
$\left\{ \begin{array}{l} \dot{\underline{x}} = \underline{A}\underline{x} + \underline{B}\underline{u} \\ \underline{y} = \underline{C}\underline{x} + \underline{D}\underline{u} \end{array} \right\} \quad \Sigma(A, B, C, D) \quad - \quad \left\{ \begin{array}{l} \underline{x}_{t+1} = \underline{\phi}\underline{x}_t + \underline{\Gamma}\underline{u}_t \\ \underline{y}_t = \underline{H}\underline{x}_t + \underline{D}\underline{u}_t \end{array} \right\} \quad \Sigma(\underline{\phi}, \underline{\Gamma}, \underline{H}, \underline{D})$ $\underline{u} \in U \subseteq \mathbb{R}^m ; \underline{x} \in X \subseteq \mathbb{R}^n ; \underline{y} \in Y \subseteq \mathbb{R}^p$		
$\mathcal{A}$ ensemble des états accessibles $\mathcal{O}$ ensemble des états observables $\mathcal{C}$ ensemble des états commandables $\mathcal{R}$ ensemble des états reconstructibles $\mathcal{A} \subseteq \mathcal{C}$ $\Sigma(A, B, C, D) / \Sigma(\underline{\phi}, \underline{\Gamma}, \underline{H}, \underline{D})$ $\Delta_t(A, B) = (B, AB, \dots, A^{t-1}B)$ $\Lambda_t(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{t-1} \end{bmatrix}$	DUALITE	$\emptyset$ $\mathcal{A}$ $\mathcal{R}$ $\mathcal{C}$ $\emptyset \subseteq \mathcal{R}$ $\Sigma^*(A^T, C^T, B^T, D^T) / \Sigma^*(\underline{\phi}^T, \underline{H}^T, \underline{\Gamma}^T, \underline{D}^T)$ $\Lambda_t^{*T}(A^T, C^T)$ $\Delta_t^{*T}(A^T, B^T)$
<u>Systèmes continus</u> $H(s) = C(sI - A)^{-1}B + D$ $\rho[\Delta_t(A, B)] = n$ $\rho[\Lambda_t(A, C)] = n$	<u>Réalisation</u> $f_\Sigma = f \Leftrightarrow U \xrightarrow{f} Y \Leftrightarrow f = h \circ g$ X $\left\{ \begin{array}{l} \text{Accessibilité } X = \mathcal{A} \\ g \text{ surjectif} \\ \text{Commandabilité : } X = \mathcal{C} \\ U \xrightarrow{g'} \begin{cases} R[\underline{\phi}^t] \\ R[e^{At}] \end{cases} \\ \text{surjectif} \\ \text{Observabilité : } X = \mathcal{O} \\ h \text{ injectif} \\ \text{Reconstructibilité } X = \mathcal{R} \\ \left. \begin{array}{l} R[\underline{\phi}^t] \\ R[e^{At}] \end{array} \right\} \xrightarrow{h'} Y \\ \text{injectif} \end{array} \right.$	<u>Systèmes discrets</u> $H(z) = H(zI - \underline{\phi})^{-1}\underline{\Gamma} + D$ $\rho[\Delta_t(\underline{\phi}, \underline{\Gamma})] = n$ $R[\Delta_t(\underline{\phi}, \underline{\Gamma})] \supseteq R[\underline{\phi}^t]$ $t \leq \bar{t}$ $\rho[\Lambda_t(\underline{\phi}, H)] = n$ $R[\Lambda_t^T] \supseteq R[\underline{\phi}^t]$ $t \leq \bar{t}$
Réalisation irréductible factorisation canonique $X \left\{ \begin{array}{l} \text{accessible} \\ \text{et} \\ \text{observable} \end{array} \right\} \quad \left\{ \begin{array}{l} g \text{ surjectif} \\ \text{et} \\ h \text{ injectif} \end{array} \right\}$		

BIBLIOGRAPHIE

- [1] C.T. CHEN "Introduction to linear system theory"
Holt, Rinehart, Winston, Inc. 1970.

- [2] R.E. KALMAN "Topics in mathematical system theory"
P.L. FALB Mc. Graw Hill Book Company, 1969
M.A. ARBIB

- [3] L.A. ZADEH "Linear system theory : the state space approach"
C.A. DESOER Mc Graw Hill Book Company, 1963

- [4] H. KWAKERNAAK "Linear optimal control systems"
R. SIVAN Wiley - Interscience, 1972

- [5] P. FAURRE "Eléments d'Automatiques"
M. DEPEYROT Dunod, 1974

- [6] P. FAURRE "Navigation inertielle optimale et filtrage statistique"
Dunod, 1971

- [7] P.R. HALMOS "Finite dimensional vector spaces"
Princeton University press, 7^e édition 1955

- [8] W.M. WONHAM "Random differential equation in control theory"
In - probabilistic method in applied mathematics
Ed. Bharucha-Reid, Academic Press, 1970.

- [9] P. DORATO "Optimal linear regulators : the discrete time case"
A.H. LEVIS I.E.E.E. tr. Aut. Cont. Vol. AC-16-6, Dec 71

- [10] H.P. ZEIGER "Ho's algorithm, commutative diagrams, and the
uniqueness of minimal linear systems.
Inf. and Control, 11, pp. 71-79, 1971.

Chapitre III

FORMES CANONIQUES ET STRUCTURE INVARIANTE

o o

o

- 1 - Introduction
- 2 - Préliminaires
- 3 - Systèmes monovariabiles
- 4 - Systèmes mono-sortie et mono-entrée
- 5 - Définition d'une structure pour l'opérateur (A,B)
- 6 - Formes canoniques multivariabiles
- 7 - Conclusion
- Bibliographie

Résumé :

L'objet de ce chapitre est de développer une paramétrisation efficace des systèmes dynamiques multivariabiles, en vue de leur réalisation (identification).

Cette paramétrisation sera caractérisée par un ensemble complet d'invariants indépendants définissant la structure du système.

Il en découlera des formes canoniques spécifiques constituant une extension des résultats classiques relatifs aux systèmes monovariabiles.

1 - INTRODUCTION

Le problème de réalisation consiste, rappelons le, à déterminer à partir d'un modèle entrées-sorties, un système dynamique Σ_n , ayant le même comportement externe. On a vu que les critères d'accessibilité et d'observabilité permettaient de caractériser la valeur minimale de n , compatible avec les données.

L'étude des formes canoniques, et donc implicitement de la structure de Σ , a pour but de poser dans les meilleures conditions le problème ultime qui consiste à déterminer les éléments de matrices A, B, C, D (ou Φ, Γ, H, D). Il est clair que cette optimisation ne peut porter que sur les matrices A, B, C (ou Φ, Γ, H), la matrice D étant indépendante de la représentation d'état ; aussi ferons nous abstraction de cette dernière au cours de ce chapitre.

Nous considérerons d'abord les systèmes monovariables, pour lesquels la détermination d'une paramétrisation optimale de Σ est relativement simple. Ce cas, intéressant en tant que référence, sera traité en détail au § 3.

On abordera ensuite les systèmes mono-sortie et leur dual mono-entrée § 4. La présence de plusieurs entrées ou plusieurs sorties mettront en évidence la nécessité de définir une structure interne caractérisant la manière dont les états sont atteints par les entrées ou observés au moyen des sorties. La définition d'un ensemble complet d'invariants indépendants viendra expliciter ce point de vue qualitatif § 5.

Ces invariants seront, en fait, les paramètres caractéristiques à partir desquels seront construits des formes canoniques multivariables spécifiques (§ 6), généralisant les différentes structures introduites dans le cas monovariable.

Pour ne pas alourdir cet exposé, on a rappelé dans un paragraphe préliminaire (§ 2) quelques résultats concernant la structure des opérateurs linéaires [1].

2 - PRELIMINAIRES

Soit A un opérateur linéaire dans R^n . Introduisons sa matrice caractéristique $(\lambda \mathbf{1} - A)$ où $\mathbf{1}$ est la matrice unité. Soit $\Delta(\lambda)$ le polynome caractéristique de A , c'est-à-dire :

$$\Delta(\lambda) \triangleq \det |\lambda \mathbf{1} - A| = \lambda^n + \alpha_1 \lambda^{n-1} + \dots + \alpha_{n-1} \lambda + \alpha_n$$

... dont les racines (valeurs propres de A) sont notées $\lambda_1, \lambda_2, \dots, \lambda_r$, $r \leq n$ respectivement de multiplicité μ_i , $i=1, \dots, r$; avec $\mu_i > 0$ et $\mu_1 + \mu_2 + \dots + \mu_r = n$

On appelle polynome annulateur de A tout polynome $\psi(\lambda)$ tel que $\psi(A) = 0$

Propriété 1 : $\Delta(A) = 0$

On appelle polynome minimal de A, le polynome annulateur de A de plus bas degré et dont le coefficient du terme du plus haut degré est 1. Soit $\phi(\lambda)$ ce polynome, de degré $m \leq n$.

Désignons par $D_p(\lambda)$ le P.G.C.D. de tous les mineurs d'ordre p de la matrice caractéristique (coefficient du terme du plus haut degré égal à 1).

Notons que $D_n(\lambda) \equiv \Delta(\lambda)$ et $D_0(\lambda) \equiv 1$

Propriété 2 :

Dans la suite $D_n(\lambda), D_{n-1}(\lambda), \dots, D_1(\lambda)$ chaque polynome est divisible par le suivant.

Définissons alors les polynomes :

$$i_1(\lambda) = \frac{D_n(\lambda)}{D_{n-1}(\lambda)} ; i_2(\lambda) = \frac{D_{n-1}(\lambda)}{D_{n-2}(\lambda)} ; \dots ; i_n(\lambda) = \frac{D_1(\lambda)}{D_0(\lambda)}$$

Les polynomes $i_p(\lambda), p = 1, \dots, n$ sont les polynomes invariants[†] de l'opérateur A.

Propriété 3 :

$$\Delta(\lambda) = i_1(\lambda) \cdot i_2(\lambda) \cdot \dots \cdot i_n(\lambda) \text{ et } i_1(\lambda) \equiv \phi(\lambda)$$

Considérons maintenant la matrice compagnon :

$$L = \begin{bmatrix} 0 & & & -\alpha_m \\ 1 & & & \vdots \\ & \ddots & & \vdots \\ & & 1 & -\alpha_1 \end{bmatrix}$$

dont le polynome caractéristique est :

$$g(\lambda) = \lambda^m + \alpha_1 \lambda^{m-1} + \dots + \alpha_m$$

[†] En raison de l'invariance de ces polynomes pour des λ -matrices équivalentes et en particulier pour des matrices semblables.

Propriété 4 : $i_1(\lambda) = g(\lambda)$; $i_2(\lambda) = \dots = i_m(\lambda) = 1$; une matrice compagne est donc caractérisée par un seul polynôme invariant identique avec son polynôme caractéristique et donc son polynôme minimal.

Par suite ...

Propriété 5 : La matrice A est semblable à une matrice diagonale bloc compagne associée à ses polynômes invariants autres que 1.

$$A^* = \begin{bmatrix} L_1 & & 0 \\ & \ddots & \\ 0 & & L_t \end{bmatrix} \quad \text{où} \quad L_p = \begin{bmatrix} 0 & & -\beta_{n_p}^p \\ 1 & & \vdots \\ & \ddots & \vdots \\ & & 1 & -\beta_1^p \end{bmatrix}; \quad i_p = \lambda^{n_p} + \beta_1^p \lambda^{n_p-1} + \dots + \beta_{n_p}^p$$

$p = 1, 2, \dots, t \leq n$

Propriété 6 : Une matrice A n'est semblable à la matrice compagne associée à son polynôme caractéristique que si, et seulement si,

$$\Delta(\lambda) = \phi(\lambda) \iff i_2(\lambda) = \dots = i_n(\lambda) = 1.$$

Explicitons ce résultant en fonction du spectre de A. Pour cela factorisons i_p soit :

$$i_p(\lambda) = (\lambda - \lambda_1)^{\delta_{p1}} (\lambda - \lambda_2)^{\delta_{p2}} \dots (\lambda - \lambda_r)^{\delta_{pr}}; \quad \delta_{pj} \geq 0 \text{ et } \delta_{pj} \geq \delta_{(p+1)j}$$

Par définition les $(\lambda - \lambda_i)^{\delta_{pi}} \neq 1$ sont les diviseurs élémentaires de A.

Introduisons maintenant la matrice de Jordan :

$$J = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

ℓ

caractérisée par le seul diviseur élémentaire $(\lambda - \lambda_i)^\ell$. Une formulation analogue de la propriété 6 sera donc :

Propriété 7 : Une condition nécessaire est suffisante pour qu'une matrice A soit semblable à la matrice compagne associée à son polynôme caractéristique est qu'à chaque valeur propre multiple ne soit associé qu'un diviseur élémentaire $(\lambda - \lambda_i)^{\ell_i}$, c'est-à-dire un seul bloc de Jordan ou encore un seul vecteur propre.

Remarque : Si la matrice A est diagonalisable tout en ayant des valeurs propres multiples, ceci signifie que son polynome minimal est différent de son polynôme caractéristique et que, par conséquent, elle ne pourra être semblable à un bloc compagnon unique, (celui de son polynôme caractéristique).

Voyons maintenant une interprétation géométrique de ces résultats.
Considérons la suite S :

$$\underline{b}, A\underline{b}, \dots, A^{p-1}\underline{b}$$

où p est tel que $A^p \underline{b}$ soit linéairement dépendant des p vecteurs précédents :

$$A^p \underline{b} = -\alpha_1 A^{p-1} \underline{b} - \alpha_2 A^{p-2} \underline{b} \dots - \alpha_p \underline{b} \quad (0 < p \leq n)$$

Soit $\Psi(\lambda)$ le polynome défini par :

$$\Psi(\lambda) = \lambda^p + \alpha_1 \lambda^{p-1} + \dots + \alpha_p$$

Alors :

$$\Psi(A)\underline{b} = \underline{0}$$

$\Psi(A)$ est appelé polynome annulateur de vecteur $\underline{b}$. Le polynome minimal de R^n est par définition $\phi(\lambda)$ le polynome minimal de A qui n'est autre que le P.P.C.M. des polynomes minimaux des vecteurs $\underline{b}$ et R^n .

Appelons I le sous espace engendré par la suite S. Ce sous espace est le sous espace cyclique[†] engendré par $\underline{b}$; $\Psi(\lambda)$ est son polynome associé.

Propriété 8 : Le sous espace cyclique I est invariant par rapport à A ;

$$\underline{x} \in I \Rightarrow A\underline{x} \in I.$$

Propriété 9 : Un espace R^n est cyclique si, et seulement si, sa dimension (n) est égale au degré de son polynome minimal. Dans la base S correspondante, la matrice A a alors la forme compagnon. On dit dans ce cas que la matrice A est cyclique.

Dans ces conditions la propriété 5 s'énoncera maintenant

Propriété 10 : Relativement à l'opérateur A, l'espace R^n peut être décomposé en une somme directe de sous espaces cycliques invariants $I_1, I_2, \dots, I_t$;
 $R^n = I_1 \oplus I_2 \oplus \dots \oplus I_t$ dont les polynomes minimaux sont les polynomes invariants (autres que 1) de la matrice A.

[†] En raison de l'aspect de S et de la définition de p.

Remarques: Les propriétés 6 et 7 expriment donc des conditions nécessaires et suffisantes pour qu'une matrice soit cyclique.

On a parlé jusqu'ici de sous espace cyclique et invariant en termes "absolu" auquel correspondaient des formes "canoniques" pour A du type bloc diagonale. On va maintenant introduire ces mêmes concepts en terme cette fois "relatif" ou "modulo" un autre sous espace.

Soit deux sous-espaces E_1 et E_2 . Nous disons que E_2 est invariant modulo E_1 si :

$$\forall \underline{x} \in E_2, A \underline{x} = \underline{x}_1 + \underline{x}_2 \text{ avec } \underline{x}_1 \in E_1, \text{ et } \underline{x}_2 \in E_2$$

On définira également :

$$\underline{x} \equiv \underline{0} \pmod{E_1} \iff \underline{x} \in E_1$$

Par ailleurs $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k$ seront linéairement indépendants (mod. E_1) si :

$$\alpha_1 \underline{x}_1 + \alpha_2 \underline{x}_2 + \dots + \alpha_k \underline{x}_k \equiv \underline{0} \pmod{E_1} \Rightarrow \alpha_i = 0, i = 1, 2, \dots, k.$$

Enfin d'une façon plus générale, pour transposer des énoncés "absolus" en énoncés "relatifs", on remplace le symbole "=" par " $\equiv \text{mod. } I$ " où I est un sous espace invariant par rapport à A.

Application : Considérons une matrice A, $n \times n$, ayant deux polynomes invariants de degré n_1 et n_2 ($n_1 + n_2 = n$). Elle est donc semblable à une matrice bloc diagonale :

$$L = \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix} \quad \text{avec } L_p, p = 1, 2 \quad L_p = \begin{bmatrix} 0 & & & -\alpha_n^p \\ 1 & & & \\ & \ddots & & \\ 0 & & 1 & -\alpha_1^p \end{bmatrix}$$

$$\text{où } i_p = \lambda^{n_p} + \alpha_1^p \lambda^{n_p-1} + \dots + \alpha_{n_p}^p$$

On fait ainsi apparaître une décomposition de R^n en une somme directe de deux sous-espaces cycliques invariants de dimension n_1 et n_2 . Il existe donc deux vecteurs de R^n , $\underline{a}$ et $\underline{b}$ tels que :

$$R^n = I_1 \oplus I_2 \text{ et } I_1 = R \left[\underline{a}, A\underline{a}, \dots, A^{n_1-1} \underline{a} \right] ; I_2 = R \left[\underline{b}, A\underline{b}, \dots, A^{n_2-1} \underline{b} \right]$$

et dont i_1 et i_2 sont les polynomes minimaux.

Considérons maintenant un vecteur $\underline{c}$ quelconque dans R^n et I'_1 l'espace cyclique engendré par $\underline{c}$, soit donc $I'_1 = R \left[\underline{c}, A\underline{c}, \dots, A^{n'_1-1} \underline{c} \right]$. Définissons un second vecteur $\underline{d} \in R^n$ mais $\notin I'_1$, désignons alors par I'_2 l'espace engendré par $[\underline{d}, A\underline{d}, \dots, A^{n'_2-1} \underline{d}]$ tel que $A^{n'_2} \underline{d}$ soit le premier vecteur linéairement dépendant de $\underline{c}, A\underline{c}, \dots, A^{n'_1-1} \underline{c}, \underline{d}, A\underline{d}, \dots, A^{n'_2-1} \underline{d}$.

Admettons pour simplifier que $n'_1 + n'_2 = n \iff R^n = I'_1 \oplus I'_2$. Il est clair que maintenant I'_2 n'est invariant et cyclique que modulo I'_1 . Soit $\phi_1(\lambda)$ le polynome minimal de I'_1 (c'est à dire de $\underline{c}$) et $\phi_2(\lambda)$ le polynome minimal de I'_2 (de $\underline{d}$) modulo I'_1 . Explicitons les dans les termes suivants :

$$\phi_1(\lambda) = \lambda^{n'_1} + \beta_1^1 \lambda^{n'_1-1} + \dots + \beta_{n'_1}^1$$

$$\phi_2(\lambda) = \lambda^{n'_2} + \beta_1^2 \lambda^{n'_2-1} + \dots + \beta_{n'_2}^2$$

Alors par définition on sait que :

$$A^{n'_2} \underline{d} = -\beta_1^2 A^{n'_2-1} \underline{d} + \dots - \beta_{n'_2}^2 \underline{d} + \gamma_1 A^{n'_1-1} \underline{c} + \dots + \gamma_{n'_1} \underline{c}$$

et dans la base $I'_1 \oplus I'_2$, A a la forme "diagonale-bloc-compagnon-couplé" :

$$L' = \begin{bmatrix} L'_1 & C \\ 0 & L'_2 \end{bmatrix} \quad \text{où} \quad C = \begin{bmatrix} \vdots & \gamma_{n'_1} \\ 0 & \vdots \\ \vdots & \gamma_1 \end{bmatrix}$$

Remarque : Ces deux formes "canoniques" ne présentent pas nécessairement la même structure, c'est à dire $n'_1 \neq n'_2$ d'une part, et, d'autre part le nombre de blocs L'_j peut être différent de celui des L_j . Notons encore que les polynomes caractéristiques des blocs L'_j , $\phi_j(\lambda)$ ne vérifient pas à priori de conditions de divisibilité.

3 - SYSTEMES MONOVARIABLES

Le problème est abordé en deux étapes. On expose d'abord comment élaborer les formes canoniques que l'on peut associer à une transmittance rationnelle. Puis étant donné un système dynamique quelconque, on verra dans quelles conditions il existe une transformation permettant d'obtenir une de ces formes.

La matrice D ne jouant ici aucun rôle, on s'en tiendra aux transmittances strictement propres, sans restreindre la généralité de ce qui suit.

Soit donc :

$$G(s) = C(s \mathbf{1} - A)^{-1} B = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \dots + b_n}{s^n + a_1 s^{n-1} + \dots + a_n} \Delta \frac{N(s)}{D(s)} \quad (\text{III-1a})$$

ou encore

$$F(z) = H(z \mathbf{1} - \Phi)^{-1} \Gamma \equiv \frac{b_1 z^{n-1} + b_2 z^{n-2} + \dots + b_n}{z^n + a_1 z^{n-1} + \dots + a_n} \Delta \frac{N(z)}{D(z)} \quad (\text{III-1b})$$

Développons maintenant ces deux expressions en série, autour de $|s| = \infty$ et $|z| = \infty$ respectivement. Il vient [2] :

$$G(s) = \sum_{i=1}^{\infty} g_i s^{-i} = \sum_{i=1}^{\infty} C A^{i-1} B s^{-i} \quad (\text{III-2a})$$

et

$$F(z) = \sum_{t=1}^{\infty} f_t z^{-t} = \sum_{t=1}^{\infty} H \Phi^{t-1} \Gamma z^{-t} \quad (\text{III-2b})$$

Les quantités $\{g_i\}$ ou $\{f_t\}$ sont ce qu'on appelle les paramètres de Markov du système. Leur signification est particulièrement intéressante dans le cas discret, puisqu'ils s'identifient alors à la séquence de pondération ou réponse impulsionnelle discrète de $F(z)$. Notons toutefois que la suite $\{g_i\}$ ne constitue pas pour autant un échantillonnage de la réponse impulsionnelle $g(t)$. On a, en fait, la relation :

$$g(t) = \sum_{i=1}^{\infty} \frac{g_i t^{i-1}}{(i-1)!} .$$

Notre propos est de développer maintenant des réalisations dites "canoniques" ne dépendant que des paramètres du modèle "externe". Les problèmes discrets et continus étant absolument identiques sur le plan formel (Cf. éq. III-1 et III-2) on raisonnera dans le cas discret en raison de l'interprétation précise de $\{f_t\}$.

D'après (III-1b), il est clair que les zéros de $D(z)$ sont également les valeurs propres de Φ (dimension $n \times n$). Par suite la paramétrisation la plus simple pour Φ consiste à choisir une des formes compagnons associées ou polynômes $D(z)$. D'après la forme explicite de $(z \mathbf{1} - A)^{-1}$ [3] il est immédiat de voir que l'on peut toujours déterminer un couple (H, Γ) tel que (III-1b) soit satisfait.

Introduisons la notation suivante pour indiquer que deux réalisations sont équivalentes :

$$(\Phi, \Gamma, H) \xrightarrow{S} (\Phi', \Gamma', H') \iff (\Phi' = S^{-1} A S ; \Gamma' = S^{-1} \Gamma ; H' = HS)$$

Théorème III-1 : Les triplets (Φ_i, Γ_i, H_i) $i = 1, \dots, n$ définissent des réalisations équivalentes de $F(z)$ dans la forme canonique observable.

$$\begin{array}{ccc}
 \Phi_1 = \begin{bmatrix} -a_1 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & & & 1 \\ -a_n & & & 0 \end{bmatrix} ; \Gamma_1 = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} & \xrightleftharpoons[L^{-1}]{L} & \Phi_2 = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \\ -a_n & & & -a_1 \end{bmatrix} ; \Gamma_2 = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \\
 \\
 H_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \quad \textcircled{1} & & H_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \quad \textcircled{2} \\
 \\
 \updownarrow P & & \updownarrow P \\
 \Phi_3 = \begin{bmatrix} 0 & & & -a_n \\ 1 & & & \\ 0 & & \ddots & \\ 0 & & 1 & -a_1 \end{bmatrix} ; \Gamma_3 = \begin{bmatrix} b_n \\ \vdots \\ b_1 \end{bmatrix} & \xrightleftharpoons[L'^{-1}]{L'} & \Phi_4 = \begin{bmatrix} -a_1 & \dots & -a_n \\ 1 & & \\ 0 & & 1 & 0 \end{bmatrix} ; \Gamma_4 = \begin{bmatrix} f_n \\ \vdots \\ f_1 \end{bmatrix} \\
 \\
 H_3 = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix} \quad \textcircled{3} & & H_4 = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix} \quad \textcircled{4}
 \end{array}$$

On a noté :

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow P^{-1} = P \quad (\text{III-3})$$

$$L = \begin{bmatrix} 1 & & & & \\ & a_1 & & & \\ & \vdots & & & \\ & & a_{n-1} & & \\ & & & a_1 & \\ & & & & 1 \end{bmatrix} \iff L^{-1} = \begin{bmatrix} 1 & & & & \\ & \alpha_1 & & & \\ & \vdots & & & \\ & & \alpha_{n-1} & & \\ & & & \alpha_1 & \\ & & & & 1 \end{bmatrix} \quad (\text{III-4})$$

$$\text{où } \alpha_i = -a_i - \sum_{k=1}^{i-1} a_k \alpha_{i-k} ; i = 2, \dots, n-1 \text{ et } \alpha_1 = -a_1 \quad (\text{III-5})$$

avec enfin :

$$L' = PLP \Rightarrow L' = \begin{bmatrix} 1 & & & & \\ & a_1 & & & \\ & & a_{n-1} & & \\ & & & a_1 & \\ & & & & 1 \\ & & 0 & & \end{bmatrix} \iff L'^{-1} = \begin{bmatrix} 1 & & & & \\ & \alpha_1 & & & \\ & & \alpha_{n-1} & & \\ & & & \alpha_1 & \\ & & & & 1 \\ & & 0 & & \end{bmatrix} \quad (\text{III-6})$$

Démonstration :

La réalisation (3) est un résultat classique [4] . Les équivalences (1) - (3) et (2) - (4) sont triviales, permutation des états. Voyons comment on détermine la matrice L. D'après la définition de F(z), équation (III-1b), on a la relation entrée-sortie :

$$y_t + a_1 y_{t-1} + \dots + a_n y_{t-n} = b_1 u_{t-1} + b_2 u_{t-2} + \dots + b_n u_{t-n} \quad (\text{III-7})$$

Par définition $y_t = f_t$ pour $u_0 = 1$ et $u_t = 0 \quad \forall t \neq 0$. Par suite, il vient :

$$\begin{aligned} f_1 &= b_1 \\ f_2 + a_1 f_1 &= b_2 \\ &\vdots \\ f_n + a_1 f_{n-1} + \dots + a_{n-1} f_1 &= b_n \end{aligned}$$

qui s'écrit encore d'après (III-4)

$$L \Gamma_2 = \Gamma_1 \iff \Gamma_2 = L^{-1} \Gamma_1 \quad (\text{III-8})$$

Par ailleurs, il est trivial de vérifier que :

$$[10 \dots 0] = [10 \dots 0] L \iff H_2 = H_1 L \quad (\text{III-9})$$

Pour démontrer que $\Phi_2 = L^{-1} \Phi_1 L$ nous utiliserons le fait que les colonnes de Φ_2 sont la représentation de celles de $\Phi_1 L$ dans la base définie par les colonnes de L (Définition du changement de base) ; d'où le résultat cherché (III-10).

$$L = \begin{bmatrix} 1 & & & & \\ & a_1 & & & \\ & & \ddots & & \\ & & & a_{n-1} & \\ & & & & 1 \end{bmatrix} \Rightarrow \Phi_1 L = \begin{bmatrix} 0 & 1 & & & \\ & a_1 & & & \\ & & \ddots & & \\ & & & a_{n-2} & a_1 \\ -a_n & 0 & & & 0 \end{bmatrix} \Rightarrow \Phi_2 = \begin{bmatrix} 0 & 1 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & & -a_1 \\ & & & -a_n & \end{bmatrix} \quad (\text{III-10})$$

Remarque : Ceci généralise le résultat développé en [5], tout en conduisant à une démonstration beaucoup plus simple.

L'équivalence (3) - (4) se déduit des autres par transitivité. On a en effet :

$$(3) \xrightarrow{P} (1) \xrightarrow{L} (2) \xrightarrow{P} (4) \iff (3) \xrightarrow{PLP} (4)$$

d'où l'on tire $L' = PLP$. La forme (III-6) se déduit quant à elle de la structure des matrices L et P définies en (III-3) et (III-4).

Il ne reste plus qu'à montrer que ces réalisations sont observables. Considérons la réalisation (2) par exemple et construisons sa matrice d'observabilité, il vient :

$$\Theta = \begin{bmatrix} H_2 & & \\ H_2 & \Phi_2 & \\ & & \ddots \\ H_2 & \Phi_2^{n-1} & \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} \Rightarrow \rho[\Theta] = n \quad (\text{III-11})$$

Les quatre réalisations considérées étant équivalentes, l'observabilité de (2) entraîne celle des autres [4].

Ce résultat étant acquis, on en déduit le théorème suivant par dualité (Cf. Ch. II) :

Théorème III-2 : Les triplets $(\Phi_i, \Gamma_{ib}, H_{ib})$ $i = 1, \dots, 4$ définissent des réalisations équivalentes de $F(z)$, dans la forme canonique accessible.

$$\Phi_4 = \begin{bmatrix} -a_1 & \cdots & -a_n \\ 1 & & \\ & 0 & \\ & & 1 & 0 \end{bmatrix}; \Gamma_{4b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \xrightleftharpoons[L']{L'^{-1}} \Phi_3 = \begin{bmatrix} 0 & & -a_n \\ 1 & 0 & \\ & \ddots & \\ 0 & 1 & -a_1 \end{bmatrix}; \Gamma_{3b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$H_{4b} = [b_1 \dots b_n] \quad (4b)$$

$$H_{3b} = [f_1 \dots f_n] \quad (3b)$$

$$\updownarrow P$$

$$\updownarrow P$$

$$\Phi_2 = \begin{bmatrix} 0 & 1 & & \\ & \ddots & 0 & \\ & & 0 & 1 \\ -a_n & \cdots & -a_1 \end{bmatrix}; \Gamma_{2b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \xrightleftharpoons[L]{L^{-1}} \Phi_1 = \begin{bmatrix} -a_1 & 1 & & \\ & \ddots & 0 & \\ & & 0 & 1 \\ -a_n & & & 0 \end{bmatrix}; \Gamma_{1b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$H_{2b} = [b_n \dots b_1] \quad (2b)$$

$$H_{1b} = [f_n \dots f_1] \quad (1b)$$

Remarque : - La réalisation "ib" correspond au bloc Φ_i de la forme observable "i". La disposition 4b, 3b, 2b, 1b permet de faire correspondre les réalisations 1, 2, 3, 4 par dualité (Exemple : $\Phi_4 = \Phi_1^T$, $\Gamma_{4b} = H_1^T$, et $H_{4b} = \Gamma_1^T$).

- On reconnaîtra en 2b, la forme bien connue dite "en variables de phase".

La question que l'on se pose maintenant est de savoir s'il existe un changement de base permettant de passer de la forme observable à la forme accessible et réciproquement. Autrement dit, ces huit formes canoniques sont-elles équivalentes ?

Une remarque s'impose. Si une telle transformation existe, cela implique nécessairement que ces réalisations soient irréductibles, puisqu'alors à la fois observables et accessibles.

Considérons, par exemple, les réalisations 3 et 3b, et définissons une transformation T par :

$$T = (\Gamma_3, \Phi_3 \Gamma_3, \dots, \Phi_3^{n-1} \Gamma_3) \quad (\text{III-12})$$

Il est clair que T est régulière si, et seulement si, (Φ_3, Γ_3) est accessible, c'est à dire (Φ_3, Γ_3, H_3) irréductible. Dans cette hypothèse, on vérifie facilement que :

$$T. \Gamma_{3b} = \Gamma_3 \quad \text{et} \quad H_{3b} = H_3 . T \quad (\text{III-13})$$

Par ailleurs, en remarquant encore une fois que :

$$\Phi_3^{-1} T = (\Phi_3 \Gamma_3, \dots, \Phi_3^n \Gamma_3)$$

... exprimé dans la base "T" redonne la matrice Φ_3 , c'est à dire $\Phi_3 = T^{-1} \Phi_3 T$; on en conclut que 3 et 3b sont des réalisations équivalentes de F(z), si, et seulement si, (Φ_3, Γ_3) est accessible, c'est à dire (Φ_3, Γ_3, H_3) irréductible.

Revenons à l'hypothèse T régulière. Elle exprime une contrainte entre les paramètres $(a_i, b_i, i = 1, \dots, n)$ qui implique, au niveau de la transmittance F(z), que les polynômes N(z) et D(z) soient premiers entre eux [6]. En conséquence :

Théorème III-3 : Les huit formes canoniques introduites successivement aux théorèmes III-1 et III-2 sont équivalentes et irréductibles si, et seulement si, la transmittance F(z) est elle même irréductible (Cf. Figure 1).

Remarque : On peut associer à une transmittance réductible d'ordre n, des réalisations de même dimension ($\geq n$) qui ne soient pas équivalentes.

Abordons maintenant le second point : étant donné un système dynamique quelconque de dimension $n(\Phi, \Gamma, H)$, peut-on déterminer un changement de base permettant de le transformer en une des formes canoniques précédentes ?

Raisonnons d'abord d'un point de vue "système". Si la réalisation n'est pas observable ou accessible, la réponse est non puisqu'il s'agit de formes canoniques soit accessibles, soit observables. Dans le cas contraire, on va voir qu'il existe une telle transformation.

Supposons la réalisation accessible, et définissons la transformation T par :

$$T = (\Gamma, \Phi\Gamma, \dots, \Phi^{n-1}\Gamma) \quad (\text{III-14})$$

Dans cette nouvelle base, il est clair que Γ devient :

$$\Gamma^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (\text{III-15})$$

Exprimons maintenant ΦT dans la base T d'où Φ^* donné par (III-16) :

$$\Phi T = [\Phi\Gamma, \dots, \Phi^n\Gamma] \Rightarrow \Phi^* = \begin{bmatrix} 0 & & -a_n \\ 1 & 0 & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \vdots & 1 & -a_1 \end{bmatrix} \quad (\text{III-16})$$

avec par définition :

$$\Phi^n\Gamma + a_1\Phi^{n-1}\Gamma + \dots + a_n\Gamma = \underline{0} \quad (\text{III-17})$$

Notons par $f_1, \dots, f_n$ les n premiers points de la séquence de pondération relative à (Φ, Γ, H) . On a donc :

$$f_i = H\Phi^{i-1}\Gamma = H^*\Phi^{*i-1}\Gamma^* \quad (\text{III-18})$$

d'où

$$H^* = [f_1, \dots, f_n] \quad (\text{III-19})$$

On reconnaît là, la forme canonique $(\Phi_3, \Gamma_{3b}, H_{3b})$, à partir de laquelle se déduisent immédiatement les 3 autres formes canoniques accessibles (au moyen des matrices P et L, Cf. (III-3) et (III-4)).

Cette approche nous paraît une alternative plus simple que les nombreuses techniques de transformation proposées pour construire en particulier la forme canonique en variable de phase $(\Phi_2, \Gamma_{2b}, H_{2b})$; [7] à [14].

On obtient bien sûr des résultats analogues en raisonnant sur les formes duales dans le cas de réalisations observables.

D'un point de vue algébrique, l'existence d'une transformation T revient à définir dans quelles conditions une matrice quelconque est semblable à sa forme compagne, ou encore revient à savoir si l'espace d'état est cyclique (interprétation géométrique).

La réponse est donnée dans les propriétés 6 et 7 § 2 ; desquelles on déduit le corollaire suivant, dans l'optique "système".

Corollaire 3-1 : - Toute réalisation (Φ, Γ, H) d'un système monovariable, dont la matrice Φ possède des valeurs propres multiples et qui de plus est diagonalisable, ne pourra être ni accessible, ni observable, quels que soient Γ et H .

Remarque : C'est le seul cas où la propriété d'accessibilité (observabilité) ne dépend que de Φ et non de la paire (Φ, Γ) , $(-\Phi, H)$. Ce résultat est par ailleurs conforme au premier raisonnement.

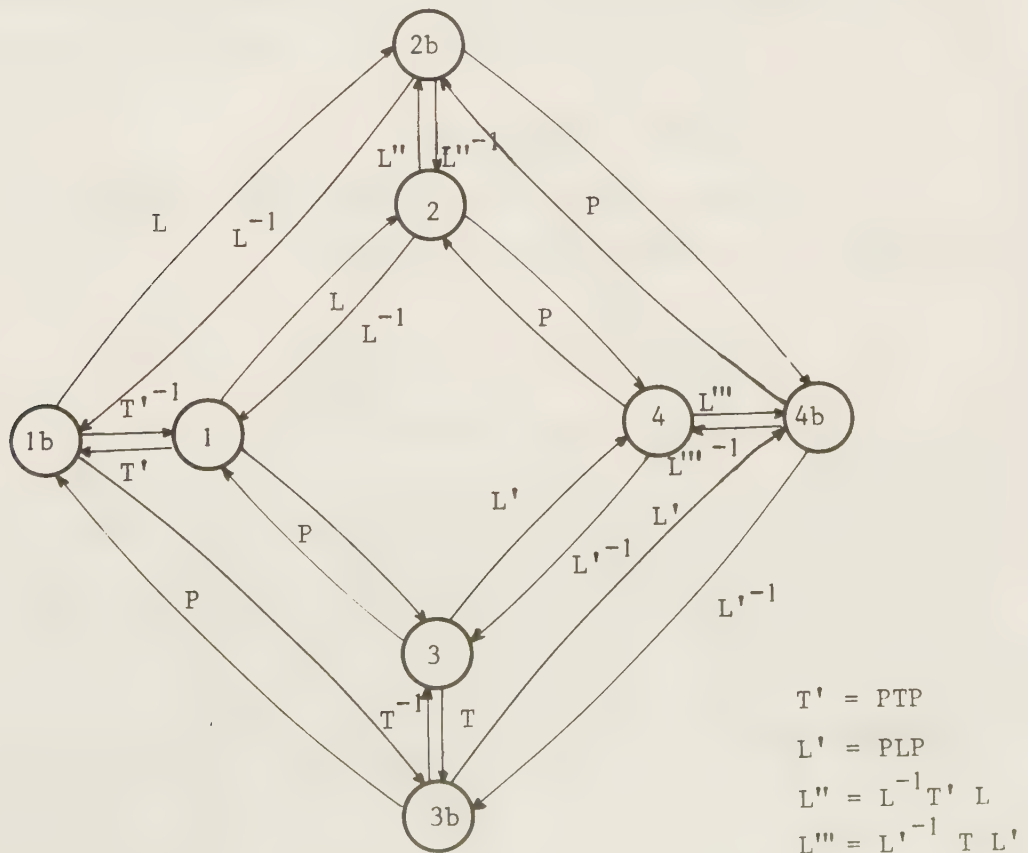


FIGURE 1

En conclusion, étant donné un système dynamique observable et accessible caractérisé par une transmittance irréductible d'ordre n , il lui correspond un ensemble de réalisations irréductibles équivalentes de dimension n . A cette classe d'équivalence on a associé deux types de paramétrisation $(a_i, b_i, i=1, \dots, n)$ ou $(a_i, f_i, i=1, \dots, n)$ conduisant chacune essentiellement à deux formes canoniques l'une observable (1)/(2), l'autre accessible (1b)/(2b).

Remarque : Nous n'avons pas introduit ici la forme canonique de Jordan, celle-ci n'étant pas directement exprimable, à partir des paramètres du modèle externe défini précédemment.

4 - SYSTEMES MONO-SORTIE et MONO-ENTREE

Ces systèmes définissent une étape naturelle entre les systèmes mono et multivariables, en ce sens qu'ils peuvent s'intégrer aux premiers (extension triviale de certaines formes canoniques), tout en préfigurant les difficultés relatives aux seconds (accessibilité de l'état au moyen de plusieurs entrées ou observabilité de l'état par l'intermédiaire de plusieurs sorties).

4-1 - Systèmes mono-sortie

On considère maintenant un système ayant m entrées et 1 sortie, caractérisé ici par sa transmittance matricielle :

$$F(z) = \left[\begin{array}{c|c} \frac{N_1(z)}{D_1(z)} & \dots & \frac{N_m(z)}{D_m(z)} \end{array} \right] \quad (\text{III-20})$$

en supposant :

$$\frac{N_i(z)}{D_i(z)} = \frac{\beta_1^i z^{n-1} + \dots + \beta_{n_i}^i}{z^{n_i} + \alpha_1^i z^{n_i-1} + \dots + \alpha_{n_i}^i}, \quad i = 1, \dots, m \quad (\text{III-21})$$

.... irréductible.

On notera :

$$\underline{f}_t^T = \left[f_t^1, \dots, f_t^m \right] \quad (\text{III-22})$$

.... la séquence de pondération relative à $F(z)$. Soit n le degré du P.P.C.M. des polynomes $D_i(z)$, $i = 1, \dots, m$. Après réduction au même dénominateur $F(z)$ s'écrira

$$F(z) = \left[\frac{\underline{b}_1^T z^{n-1} + \underline{b}_2^T z^{n-2} + \dots + \underline{b}_n^T}{z^n + a_1 z^{n-1} + \dots + a_n} \right] \quad (\text{III-23})$$

où

$$\frac{N_i(z)}{D_i(z)} = \frac{b_1^i z^{n-1} + \dots + b_n^i}{z^n + a_1 z^{n-1} + \dots + a_n} \quad \text{et} \quad \underline{b}_i^T = [b_1^i, \dots, b_m^i] \quad (\text{III-24})$$

On démontre alors, d'une manière analogue au théorème III-1, le résultat suivant :

Théorème III-4 : Les triplets (Φ_i, Γ_i, H_i) $i = 1, \dots, 4$ définissent des réalisations équivalentes de $F(z)$, dans la forme canonique observable, où les matrices Φ_i et H_i , $i = 1, \dots, 4$ sont celles du théorème III-1, et où Γ_i est donné par :

$$\Gamma_1 = \begin{bmatrix} \underline{b}_1^T \\ \vdots \\ \underline{b}_n^T \end{bmatrix} ; \quad \Gamma_2 = \begin{bmatrix} \underline{f}_1^T \\ \vdots \\ \underline{f}_n^T \end{bmatrix} ; \quad \Gamma_3 = \begin{bmatrix} \underline{b}_n^T \\ \vdots \\ \underline{b}_1^T \end{bmatrix} ; \quad \Gamma_4 = \begin{bmatrix} \underline{f}_n^T \\ \vdots \\ \underline{f}_1^T \end{bmatrix} \quad (\text{III-25})$$

A la différence du problème monovariable, il n'est plus possible de déduire de ce résultat (par dualité) une forme canonique accessible. Cet argument ne permettra que de transposer les résultats obtenus pour les systèmes mono-sortie, aux systèmes mono-entrée.

Rappelons néanmoins que dans les hypothèses adoptées (transmittances irréductibles) et compte tenu de la définition de n (degré du P.P.C.M. des $D_i(z)$, $i = 1, \dots, m$) on démontre que les réalisations canoniques du théorème III-4 sont également accessibles et par suite irréductibles ; Cf. [6] ou [3].

La difficulté rencontrée dans l'élaboration d'une forme accessible provient de la multiplicité des entrées. La transposition directe du cas monovarié conduit, par exemple, à une matrice Γ_{4b} (Cf. théorème III-2).

$$\Gamma_{4b} = \begin{bmatrix} 1 & x & \dots & x \\ 0 & x & & \\ \vdots & \vdots & & \vdots \\ 0 & x & & x \end{bmatrix} \quad (\text{III-26})$$

où x désigne un élément arbitraire. Ceci signifierait que l'espace d'état est accessible par une seule entrée (ici à priori la première), restriction injustifiable.

D'une façon générale l'accessibilité des états est partagée entre les différentes entrées, et la manière, dont elle s'effectue, définit ce qu'on appelle la structure de la paire (Φ, Γ) . Dans un contexte dual, le même problème se présente pour l'observabilité des systèmes mono-entrée, et plus généralement se trouvera simultanément posé pour les systèmes multivariés. Ceci explique les développements entrepris au § 5.

4-2 - Systèmes mono-entrée

Soit donc un système ayant cette fois une entrée et p sorties dont la transmittance est :

$$F(z) = \begin{bmatrix} \frac{N_1(z)}{D_1(z)} \\ \vdots \\ \frac{N_p(z)}{D_p(z)} \end{bmatrix} \quad (\text{III-27})$$

En conservant les hypothèses et notations précédentes, $F(z)$ s'écrira encore :

$$F(z) = \left[\frac{\underline{b}_1 z^{n-1} + \underline{b}_2 z^{n-2} + \dots + \underline{b}_n}{z^n + a_1 z^{n-1} + \dots + a_n} \right] \quad (\text{III-28})$$

avec

$$\underline{b}_i^T = \begin{bmatrix} b_1^i, \dots, b_p^i \end{bmatrix} \quad ; \quad i = 1, \dots, n.$$

Par dualité du théorème III-4 et en se référant au théorème III-2, on en déduit le ...

Théorème III-5 : Les triplets $(\phi_i, \Gamma_{ib}, H_{ib})$, $i = 1, \dots, 4$ définissent des réalisations équivalentes de $F(z)$ dans la forme canonique accessible, où les matrices ϕ_i et Γ_{ib} , $i = 1, \dots, 4$ sont celles du théorème III-2 et où H_{ib} est donnée par :

$$H_{4b} = [\underline{b}_1, \dots, \underline{b}_n] \quad ; \quad H_{3b} = [\underline{f}_1, \dots, \underline{f}_n]^\dagger \quad ; \quad H_{2b} = [\underline{b}_n, \dots, \underline{b}_1] \quad ;$$

$$H_{1b} = [\underline{f}_n, \dots, \underline{f}_1]$$

Remarque : Ces réalisations sont également observables et donc irréductibles avec les hypothèses adoptées.

Le problème qui consiste à transformer une réalisation quelconque d'un système mono-entrée ou mono-sortie en l'une de ces formes canoniques se pose en des termes voisins du cas monovarié.

Considérons le cas mono-entrée. S'il s'agit d'une réalisation accessible, la transformation T définie par :

$$T = \begin{bmatrix} \Gamma, & \Phi\Gamma, & \dots, & \Phi^{n-1} \Gamma \end{bmatrix} \quad \text{(III-29)}$$

conduit à la forme (3b). (démonstration analogue au cas monovarié).

Notons ici que si Φ est diagonalisable tout en ayant des valeurs propres multiples, le système sera non accessible quelque soit Γ ; mais à la différence du cas monovarié, il pourra néanmoins être observable.

Par dualité, la transformation T définie par :

$$T = \begin{bmatrix} H \\ H & \Phi \\ \vdots & \vdots \\ H & \Phi^{n-1} \end{bmatrix} \quad \text{(III-30)}$$

permet d'exprimer une réalisation quelconque d'un système mono-sortie observable dans la forme canonique (2) (duale de (3b)). La remarque concernant le spectre de Φ se transpose bien sûr ici.

$\dagger \underline{f}$ est ici un vecteur de dimension p .

5 - DEFINITION D'UNE STRUCTURE POUR L'OPERATEUR (A,B)

La détermination d'une forme canonique (multivariable) est basée, dans de nombreux cas, sur la sélection d'une base de l'espace d'état extraite de la matrice d'accessibilité [15], [16] ou d'observabilité [17], [18]. Pour s'assurer de l'unicité de la solution, certains auteurs [19], [20] se sont placés dans des conditions leur permettant d'arriver à une forme pseudo-compagnon pour Φ .

Dans le cas général, la multiplicité des solutions possibles a motivé une autre approche relevant plus d'un esprit différent que d'une méthodologie profondément nouvelle [21], [22].

Cette voie se situe au niveau de l'étude des caractéristiques, non pas d'une réalisation particulière mais d'une classe d'équivalence, c'est à dire s'intéresse finalement aux structures possibles d'un opérateur, ici linéaire. Elle présente par ailleurs des aspects très positifs dans le domaine de la commande, [3], [23], [21].

5-1 - Définition d'un système complet d'invariants indépendants

Soient A et B deux matrices $n \times n$ et $n \times m$ respectivement. Nous admettons que la paire (A,B) vérifie la relation :

$$\rho [\Delta_n (A, B)] = \rho [B, AB, \dots, A^{n-1} B] = n \quad (\text{III-31})$$

Cette hypothèse n'est nullement restrictive et sera levée un peu plus loin en faisant appel au théorème de décomposition (II-3). Ceci évite en particulier de compliquer, par une approche directe, des notations déjà lourdes (Cf. [16] par exemple).

Notons encore :

$$B = [\underline{b}_1, \underline{b}_2, \dots, \underline{b}_m]$$

de sorte que $\Delta_n(A, B)$ s'écrit aussi :

$$\Delta_n(A, B) = [\underline{b}_1, \dots, \underline{b}_m, A\underline{b}_1, \dots, A\underline{b}_m, A^2\underline{b}_1, \dots, A^2\underline{b}_m, \dots, A^{n-1}\underline{b}_1, \dots, A^{n-1}\underline{b}_m] \quad (\text{III-32})$$

Proposition III-2 : L'ensemble des n_i et α_{ijk} définis en (III-33) et (III-34) sont invariants par rapport à un changement de base.

En effet soit (A^*, B^*) une paire équivalente à (A, B) , c'est à dire telle que :

$$A^* = S^{-1} A S \quad \text{et} \quad B^* = S^{-1} B \quad (\text{III-35})$$

Remarquons que :

$$(A^*)^k \underline{b}_j^* = S^{-1} A^k S S^{-1} \underline{b}_j = S^{-1} A^k \underline{b}_j$$

Donc en multipliant les deux membres de (III-34) par S^{-1} , il vient :

$$(A^*)^i \underline{b}_i^* = \sum_{j=1}^i \sum_{k=1}^{n_j} \alpha_{ijk} (A^*)^{k-1} \underline{b}_j^* ; i = 1, 2, \dots, m$$

ce qui prouve que n_i et α_{ijk} sont invariants.

On dira, par la suite, que l'ensemble $(n_i, i = 1, \dots, m)$, $(\alpha_{ijk}, i = 1, \dots, m ; j = 1, \dots, i ; k = 1, \dots, n_j)$ représentent des invariants pour la paire (A, B) .

Nous allons montrer maintenant que ces invariants sont "indépendants", c'est à dire qu'étant donné un ensemble $(n_i, i = 1, m)$ et $(\alpha_{ijk}, i = 1, \dots, m ; j = 1, \dots, i ; k = 1, \dots, n_j)$ il existe une paire (A, B) pour lesquels ils répondent aux définitions (III-33) et (III-34).

Pour y parvenir, nous énoncerons le résultat intermédiaire suivant :

Proposition III - 3 : Les matrices $A^* = T^{-1} A T$ et $B^* = T^{-1} B$ ont la structure suivante :

$$A^* = \begin{bmatrix} \begin{array}{c|c} \begin{array}{ccc} \alpha_{1,1,1} & \dots & \alpha_{1,1,n_1} \\ \vdots & \ddots & \vdots \\ \alpha_{1,1,n_1} & \dots & \alpha_{1,1,n_1} \end{array} & \begin{array}{ccc} \alpha_{2,1,1} & \dots & \alpha_{2,1,n_1} \\ \vdots & \ddots & \vdots \\ \alpha_{2,1,n_1} & \dots & \alpha_{2,1,n_1} \end{array} \\ \hline \begin{array}{ccc} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{array} & \begin{array}{ccc} \alpha_{2,2,1} & \dots & \alpha_{2,2,n_2} \\ \vdots & \ddots & \vdots \\ \alpha_{2,2,n_2} & \dots & \alpha_{2,2,n_2} \end{array} \end{array} & \begin{array}{c} n_1 \\ n_2 \\ \vdots \\ n_m \end{array} \\ \hline \begin{array}{ccc} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{array} & \begin{array}{ccc} \alpha_{m,1,1} & \dots & \alpha_{m,1,n_1} \\ \vdots & \ddots & \vdots \\ \alpha_{m,2,n_2} & \dots & \alpha_{m,2,n_2} \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{array} \end{array} \quad \begin{array}{c} n \\ n \\ \vdots \\ n \end{array} \\ \hline \begin{array}{ccc} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{array} & \begin{array}{ccc} \alpha_{m,m,1} & \dots & \alpha_{m,m,n_m} \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{array} \end{array} \quad \begin{array}{c} n_1 \\ n_2 \\ \vdots \\ n_m \end{array}$$

(III-36)

Démonstration :

La relation $B^* = T^{-1} B \iff T B^* = B$ or cette dernière est triviale à vérifier.

Pour A^* raisonnons en termes de changement de base et pour cela explicitons AT , il vient :

$$AT = \left[Ab_{\underline{1}}, \dots, A^{n_1}_{\underline{b_1}}, Ab_{\underline{2}}, \dots, A^{n_2}_{\underline{b_2}}, \dots, Ab_{\underline{m}}, \dots, A^{n_m}_{\underline{b_m}} \right] \quad (III-37)$$

Compte-tenu de (III-34), les colonnes de AT exprimées dans la base T donnent bien A^* .

Proposition III-4 : Les invariants n_i, α_{ijk} définis par (III-33) et (III-34) sont indépendants.

Démonstration : Soit une paire (A, B) dont les invariants n_i et α_{ijk} satisfont (III-33) et (III-34). D'après la proposition III-2, ils sont les invariants de la paire (A^*, B^*) . Par conséquent, étant donné un ensemble de paramètres $(n_i, i = 1, \dots, m)$ et $(\alpha_{ijk}, i = 1, \dots, m; j = 1, \dots, i; k = 1, \dots, n_j)$, la paire (A^*, B^*) construite à partir de ceux-ci (III-36) répond à la question.

Proposition III-5 : Les paramètres n_i et α_{ijk} définissent un ensemble "complet" d'invariants ; c'est à dire si deux paires (A,B) et (A',B') de même dimension ont mêmes invariants, elles sont alors équivalentes.

Démonstration : D'après ce qui précède il existe une matrice T de la forme (III-33), telle que :

$$A^* = T^{-1} A T \quad \text{et} \quad B^* = T^{-1} B$$

Par hypothèse, il existera également une matrice T' ayant même structure que T et telle que :

$$A^* = T'^{-1} A' T' \quad \text{et} \quad B^* = T'^{-1} B'$$

de sorte que :

$$T^{-1} B = T'^{-1} B' \implies B = T T'^{-1} B' \triangleq R^{-1} B' \quad (\text{III-38})$$

de même :

$$T'^{-1} A' T' = T^{-1} A T \implies A = T T'^{-1} A' T' T^{-1} \triangleq R^{-1} A' R \quad (\text{III-39})$$

De (III-38) et (III-39) découle l'équivalence de (A,B) et (A', B') .

Nous résumerons en énonçant le ...

Théorème III-6 : Etant donnée une paire (A,B) satisfaisant (III-31), les paramètres $(n_i, i=1, \dots, m)$ et $(\alpha_{ijk}, i=1, \dots, m ; j=1, \dots, i ; k=1, \dots, n_j)$ répondant aux définitions (III-33) et (III-34) constituent un ensemble complet d'invariants indépendants vis à vis de l'équivalence $A' = S^{-1} A S, B' = S^{-1} B$.

Remarque : Ce théorème reste valable si un terme $\underline{b}_i$ n'intervient pas dans T c'est à dire si $n_i = 0$. Ceci signifie que :

$$\underline{b}_i = \sum_{j=1}^{i-1} \sum_{k=1}^{n_j} \alpha_{ijk} A^{k-1} \underline{b}_j \quad (\text{III-40})$$

où les α_{ijk} sont les coefficients exprimant la dépendance linéaire de $\underline{b}_i$ en fonction des vecteurs déjà sélectionnés (donc situés à sa gauche).

La structure de (A, B) définie par sa forme dite canonique (A^*, B^*) se trouve modifiée de la façon suivante : le bloc compagnon numéro i de dimension $n_i \times n_i$ disparaît et les paramètres α_{ijk} intervenant dans (III-40) normalement situés au dessus de ce bloc compagnon dans A^* (termes dits de couplage) se retrouvent dans les $(n_1 + n_2 + \dots + n_{i-1})$ premières positions

de la colonne i de B^* .

Exemple : $n_1 = 3, n_2 = 0, n_3 = 2$

$$A^* = \left[\begin{array}{ccc|cc} 0 & 0 & \alpha_{1,1,1} & 0 & \alpha_{3,1,1} \\ 1 & 0 & \alpha_{1,1,2} & 0 & \alpha_{3,1,2} \\ 0 & 1 & \alpha_{1,1,3} & 0 & \alpha_{3,1,3} \\ \hline & \bigcirc & & 0 & \alpha_{3,3,1} \\ & & & 1 & \alpha_{3,3,2} \end{array} \right] ; B^* = \left[\begin{array}{ccc} 1 & \alpha_{2,1,1} & 0 \\ 0 & \alpha_{2,1,2} & 0 \\ 0 & \alpha_{2,1,3} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right]$$

On peut interpréter l'ensemble de ces résultats de la manière suivante. Les invariants définis en (III-33) et (III-34) jouent un double rôle. Ils définissent d'une part, par l'intermédiaire des n_i , la structure de l'opérateur (A,B) , la paire (A,B) (en tant que matrice) en étant l'expression dans une certaine base de l'espace d'état X , (engendré par Δ_n (III-32)). D'autre part ils caractérisent au moyen des α_{ijk} la paramétrisation de cet opérateur à travers une forme canonique de la classe d'équivalence (A,B) .

Il faut remarquer que cette paramétrisation est réalisée avec un nombre minimum de degrés de liberté. Notons n_α le nombre de α_{ijk} . D'après (III-34), il vient :

$$n_\alpha = \underbrace{n_1 + (n_1 + n_2) + \dots + (n_1 + n_2 + \dots + n_m)}_{m \text{ termes}} \quad (\text{III-41})$$

$$\Rightarrow n_\alpha = nm - \sum_{i=2}^m (i-1) n_i \leq nm$$

Ce nombre de paramètres n_α est bien évidemment caractéristique de la définition de la matrice T (III-33). Il est clair qu'en choisissant de sélectionner les vecteurs $\underline{b}_i$ dans un ordre différent de l'ordre naturel $1, 2, \dots, m$, conduirait à d'autres valeurs pour n_α . Par suite, il existe nécessairement une permutation appropriée des vecteurs colonnes de B pour laquelle n_α est minimum. On vérifie facilement d'après (III-41) que ce minimum est atteint en sélectionnant, en premier, le vecteur $\underline{b}_i$ qui correspondra au n_1 le plus petit, puis en retenant un second vecteur $\underline{b}_j$ tel que n_2 soit minimum, et ainsi de suite. On notera, par ailleurs, que le nombre maximum de paramètres ($n_\alpha = n \cdot m$) est obtenu lorsque la paire (A,B) a la structure $(n, 0, \dots, 0)$ quel que soit le vecteur $\underline{b}_i$ sélectionné en premier, c'est à dire quand les m vecteurs $\underline{b}_i$ engendrent un espace cyclique de dimension n . La matrice A^* est alors du type compagnon.

Poursuivons maintenant en s'intéressant d'un point de vue "système" à la signification précise de la structure de l'opérateur (A, B) défini dans la forme canonique (A^*, B^*) .

Admettons pour l'instant que $m = 2$ et envisageons le problème sous l'angle d'accessibilité (par exemple). Désignons par $n_\alpha + n_\beta$ le nombre d'états accessibles par la première entrée u^1 , donc :

$$\rho [\underline{b}_1, A\underline{b}_1, \dots, A^{n-1}\underline{b}_1] = n_\alpha + n_\beta \quad (\text{III-42})$$

... et par $n_\beta + n_\gamma$ le nombre d'états accessibles par la seconde entrée, u^2 , soit :

$$\rho [\underline{b}_2, A\underline{b}_2, \dots, A^{n-1}\underline{b}_2] = n_\beta + n_\gamma \quad (\text{III-43})$$

En notant $B = (\underline{b}_1, \underline{b}_2)$, on a bien sûr :

$$\rho [B, AB, \dots, A^{n-1}B] = n_\alpha + n_\beta + n_\gamma \quad (\text{III-44})$$

En appelant α, β, γ les sous espaces de X de dimension respectivement n_α, n_β et n_γ , on en déduit que $\alpha \oplus \beta$ est accessible par u^1 , et $\beta \oplus \gamma$ est accessible par u^2 , et que γ n'est accessible que de u^2 (Cf. Fig. 2).

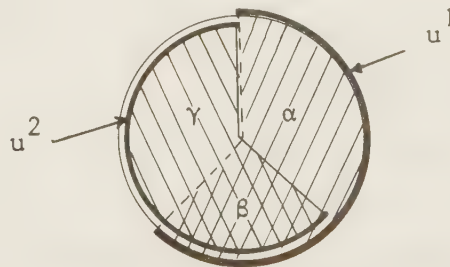


Figure 2

Compte tenu de la définition de n_1 et de (III-42), on a $n_1 = n_\alpha + n_\beta$. De même de (III-43), il vient $n_2 = n_\beta + n_\gamma$. C'est l'existence de l'espace β ($n_\beta \neq 0$) qui se traduit par la présence d'une colonne de couplage dans A^* . Notons encore que $n_\beta \neq 0$ implique que le sous espace γ ne soit cyclique et invariant modulo $\alpha \oplus \beta$. "Couplage" et "modulo" sont synonymes.

Plus généralement les sous espaces X_i de dimension n_i ne sont accessibles que par l'entrée u^i et leur somme directe est accessible par les i premières entrées. Ces sous espaces X_i sont également cycliques et invariants modulo $X_1 \oplus X_2 \oplus \dots \oplus X_{i-1}$, en raison toujours de la colonne de couplage dans A^* (et réciproquement).

Dans certains cas, il est intéressant de disposer d'une forme complètement découplée. Reprenons pour commencer l'exemple déjà considéré où $m = 2$. Eliminer le couplage revient donc à faire disparaître le sous espace β . Un moyen pour y parvenir consiste à s'assurer la divisibilité successive des polynomes caractéristiques des blocs compagnons ; ou ce qui revient au même à définir successivement des sous espaces cycliques invariants par rapport à la somme directe de ceux déjà obtenus. En se référant à l'application développée au § 2, il s'agit de déterminer des vecteurs engendrant ces sous espaces cycliques. D'après ce qui précède il n'est pas toujours possible de les sélectionner parmi les colonnes de B. Vogt et Gunderson [24] ont développé une technique pour construire ces vecteurs, mais difficile à mettre en oeuvre et qui globalement ne correspond pas exactement à un changement de base.

On propose, à cet effet, une autre approche, complètement différente, ne faisant appel à aucune condition de divisibilité, et caractérisée par un changement de base dont la matrice de passage bénéficie de propriétés spécifiques.

5-2 - Découplage de l'opérateur (A,B)

Nous allons d'abord établir un résultat préliminaire concernant le cas où $m = 2$. Posons alors :

$$A^* = \begin{bmatrix} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{bmatrix} \begin{matrix} \downarrow n_1 \\ \downarrow n_2 \end{matrix} \quad \text{et} \quad S = \begin{bmatrix} \mathbb{1} & S_{12} \\ \hline 0 & \mathbb{1} \end{bmatrix} \begin{matrix} \downarrow n_1 \\ \downarrow n_2 \end{matrix} \quad (\text{III-45})$$

$\begin{matrix} \leftarrow n_1 & \leftarrow n_2 \end{matrix}$

où $\mathbb{1}$ est la matrice unité de dimension adéquate. On vérifiera immédiatement que :

$$S^{-1} = \begin{bmatrix} \mathbb{1} & -S_{12} \\ \hline 0 & \mathbb{1} \end{bmatrix} \quad (\text{III-46})$$

On effectue alors le changement de base défini par S, soit :

$$\tilde{A} = S^{-1} A S = \begin{bmatrix} A_{11} & A_{11} S_{12} - S_{12} A_{22} + A_{12} \\ \hline 0 & A_{22} \end{bmatrix}$$

La matrice S , c'est à dire S_{12} , est alors déterminée de telle sorte que :

$$S_{12} A_{22} - A_{11} S_{12} = A_{12} \quad (\text{III-47})$$

équation dont la solution peut être obtenue par l'algorithme de Bartel et Stewart [25] dont une variante tenant compte de la forme compagnon de A_{11} et A_{22} serait particulièrement efficace. Sur le plan théorique l'existence et l'unicité de la solution S_{12} n'est garantie que si (et seulement si) les valeurs propres de A_{11} et A_{22} sont différentes, ce qui est généralement le cas en présence d'un couplage (ne serait-ce que numériquement !).

Appliquons maintenant cette transformation à la matrice B^* , il vient :

$$\tilde{B} = S^{-1} B = \begin{bmatrix} 1 & & & c_1 \\ 0 & & & \vdots \\ & & & 0 \\ \hline & & 1 & \\ 0 & & 0 & \vdots \\ & & & 0 \end{bmatrix} \quad \begin{matrix} \updownarrow n_1 \\ \updownarrow n_2 \end{matrix} \quad (\text{III-48})$$

avec $S_{12} = \begin{bmatrix} c_1 & c_2 & \dots & c_{n_2} \end{bmatrix} \quad (\text{III-49})$

En résumé la méthode proposée élimine, par un changement de base approprié, le couplage de la matrice A^* pour le reporter dans B^* , ce dernier n'étant autre que la première colonne de la matrice S_{12} .

La généralisation pour $m > 2$ s'effectue de la manière suivante. Notons maintenant :

$$A^* = \begin{bmatrix} A_{11}^* & A_{12}^* & A_{1m}^* \\ & A_{22}^* & A_{2m}^* \\ & & \vdots \\ & & A_{mm}^* \end{bmatrix} \quad ; \quad S_1 = \begin{bmatrix} & S_{1m} \\ & S_{2m} \\ & \vdots \\ & S_{m-1,m} \\ \hline 0 & 1 \end{bmatrix} \quad \begin{matrix} \updownarrow s_{m-1} \\ \updownarrow n_m \end{matrix}$$

(III-50)

$$\text{avec } s_j \triangleq \sum_{i=1}^j n_i \text{ et } s_0 = 0 \quad (\text{III-51})$$

La matrice S_1 est alors déterminée de telle sorte que :

$$S_1^{-1} A^* S_1 = A_1^* = \begin{bmatrix} A_{11}^* & A_{12}^* & & 0 \\ & A_{22}^* & & \\ & & \ddots & \\ & & & A_{mm}^* \end{bmatrix} \quad (\text{III-52})$$

Ce qui conduit aux équations :

$$\left. \begin{aligned} A_{11}^* S_{1m} + A_{12}^* S_{2m} + \dots + A_{1,m-1}^* S_{m-1,m} + A_{1m}^* - S_{1m} A_{mm}^* &= 0 \\ A_{22}^* S_{2m} + \dots + A_{2,m-1}^* S_{m-1,m} + A_{2m}^* - S_{2m} A_{mm}^* &= 0 \\ &\vdots \\ A_{m-1,m-1}^* S_{m-1,m} + A_{m-1,m}^* - S_{m-1,m} A_{mm}^* &= 0 \end{aligned} \right\} \quad (\text{III-53})$$

... que l'on résoud à partir de la dernière. On est ramené à un processus de substitution où les inconnues S_{im} , $i = m-1, \dots, 1$ sont déduites par résolution d'une équation linéaire du type (III-47).

On détermine ensuite une matrice S_2 de la même manière, telle que :

$$A_2^* = S_2^{-1} A_1^* S_2 = \begin{bmatrix} A_{11}^* & A_{11}^* & & 0 \\ & A_{22}^* & & \\ & & \ddots & \\ & & & A_{m-1,m-1}^* \\ & & & & A_{mm}^* \end{bmatrix} \quad (\text{III-54})$$

$$S_2 = \begin{bmatrix} 1 & S_{1,m-1} & & 0 \\ & \vdots & & \\ & & S_{m-2,m-1} & \\ 0 & & & 1 \end{bmatrix} \begin{matrix} \uparrow \\ n \\ \downarrow \end{matrix}$$

$\xrightarrow{s_{m-2}} \quad \xrightarrow{n_{m-1}} \quad \xrightarrow{n_m}$

et ainsi de suite, ce qui correspond à une transformation globale :

$$S = S_1 S_2 \dots S_{m-1} \quad (\text{III-55})$$

dont l'élément S_i est :

$$S_i = \begin{bmatrix} \overset{s_{m-i}}{\updownarrow} & \begin{array}{c|c|c} & S_{1,m-i+1} & \\ \hline \mathbf{1} & & 0 \\ \hline & S_{m-i,m-i+1} & \\ \hline & & \mathbf{1} \end{array} & \overset{n}{\updownarrow} \\ \hline & & \end{bmatrix} \quad (\text{III-56})$$

$\xleftarrow{\quad n \quad}$

... et qui conduit à la forme canonique découplée $(\tilde{A}, \tilde{B})$ défini par :

$$\tilde{A} = S^{-1} A^* S = \text{diag}(A^*) = \begin{bmatrix} A_{11}^* & & 0 \\ & A_{22}^* & \\ 0 & & A_{mm}^* \end{bmatrix}; \quad \tilde{B} = S^{-1} B = \begin{bmatrix} 1 & & & \\ 0 & c_2 & & \\ \vdots & \vdots & \ddots & \\ 0 & 1 & \dots & c_m \\ & 0 & \dots & \\ & \vdots & & \\ 0 & & & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{bmatrix} \quad (\text{III-57})$$

où c_{i+1} désigne les s_i premiers termes de la colonne $s_i + 1$ de la matrice S^{-1} .

Remarques :

- Les formes canoniques (A^*, B^*) et $(\tilde{A}, \tilde{B})$ comportent le même nombre de paramètres, certains invariants (couplage) étant reportés sous forme de combinaisons linéaires dans $\tilde{B}$. Il est intéressant de noter que la matrice $\tilde{A}$ ne dépend plus que de n paramètres.

- Dans le cas où un n_i est égal à 0 la transformation correspondante est $S_{m+1-i} = \mathbf{1}$ - Cf. (III-56).

5-3 - Généralisation au cas où $\rho [\Delta_n(A,B)] < n$

La différence essentielle avec l'hypothèse précédente réside dans la proposition III-1 qui devient maintenant :

Proposition III-6 : $n_1 + n_2 + \dots + n_m = \rho [\Delta_n(A,B)] = n' < n$

Notons également que la matrice T définie par (III-33) est maintenant de dimension $n \times n'$. Désignons alors par T' la matrice dont les n' premières colonnes sont formées par T et dont les $n - n'$, dernières sont choisies arbitrairement pourvu qu'elles satisfassent la condition T' régulière.

Effectuons le changement de base relatif à T' , soit :

$$A^{*'} = T'^{-1} A T \quad ; \quad B^{*'} = T'^{-1} B \quad (\text{III-58})$$

Appliquons alors le théorème II-3 à la paire (A, B) , compte tenu de la définition de T' , il en découle la décomposition :

$$A^{*'} = \begin{array}{c} \begin{array}{|c|} \hline \xrightarrow{n'} \\ \hline \end{array} \begin{array}{|c|} \hline A^{*'} \\ \hline \end{array} \begin{array}{|c|} \hline \xrightarrow{n} \\ \hline \end{array} \begin{array}{|c|} \hline A_{12}^{*'} \\ \hline \end{array} \\ \hline \begin{array}{|c|} \hline 0 \\ \hline \end{array} \begin{array}{|c|} \hline A_{22}^{*'} \\ \hline \end{array} \end{array} \quad ; \quad B^{*'} = \begin{array}{c} \begin{array}{|c|} \hline \xrightarrow{n'} \\ \hline \end{array} \begin{array}{|c|} \hline B^{*'} \\ \hline \end{array} \\ \hline \begin{array}{|c|} \hline 0 \\ \hline \end{array} \end{array} \quad n \quad (\text{III-59})$$

telle que :

$$\rho [\Delta_{n'}(A^{*'}, B^{*'})] = n' \quad (\text{III-60})$$

Maintenant, en raison de la forme particulière de T' et d'après les résultats acquis § 5-1, il est clair que la paire $(A^{*'}, B^{*'})$ a la structure définie en (III-36).

En conséquence, la définition des n_i , introduite par la relation (III-33) permet dans tous les cas de connaître la structure de l'opérateur (A,B) .

Ce résultat particulièrement important sera longuement exploité Ch. IV au cours de l'élaboration d'algorithmes de réalisation conduisant aux formes canoniques minimales que l'on développe maintenant.

Remarque : Les paramètres n_i, α_{ijk} représentent toujours un ensemble d'invariants indépendants, mais il n'est plus complet à moins qu'on ne se restreigne au sous système de dimension n' .

6 - FORMES CANONIQUES MULTIVARIABLES ASSOCIEES A LA STRUCTURE PRECEDENTE

Rappelons, pour mémoire, que les huit formes canoniques introduites § 3 pour les systèmes monovariabiles se généralisent d'une manière triviale, pour le problème multivariable. Soit $H(z)$ la matrice de transfert du système dont les éléments sont supposés irréductibles. Notons r le degré du P.P.C.M. de tous les dénominateurs, alors $H(z)$ peut se mettre sous la forme :

$$H(z) = \frac{B_1 z^{r-1} + B_2 z^{r-2} + \dots + B_r}{z^r + a_1 z^{r-1} + \dots + a_n} \quad (\text{III-61})$$

où les B_i ; $i = 1, \dots, r$ sont des matrices de dimension $p \times m$. On vérifie alors facilement que les théorèmes III-1 et III-2 sont encore vrais en remplaçant :

- pour le théorème III-1

$$\begin{cases} f_t \text{ par } F_t & \text{matrice } p \times m \\ b_i \text{ par } B_i & \text{matrice } p \times m \\ a_j \text{ par } a_j \mathbf{1}_p & \mathbf{1}_p \text{ matrice unité } p \times p \\ 0 \text{ par } 0_p & \text{matrice nulle } p \times p \end{cases}$$

ce qui conduit à une réalisation observable de dimension $r + p$

- pour le théorème III-2

$$\begin{cases} f_t \text{ par } F_t \\ b_i \text{ par } B_i \\ a_j \text{ par } a_j \mathbf{1}_m \\ 0 \text{ par } 0_m \end{cases}$$

ce qui conduit, dans ce cas, à une réalisation accessible de dimension $r \times m$.

Ces structures bien que simples ne présentent qu'un intérêt théorique. Il est évident que les huit formes que l'on peut construire ne sont pas semblables, ne serait-ce que parce qu'elles sont de dimensions différentes ($r \times p$ et $r \times m$) et à priori réductibles ce qui ne leur confère qu'un attrait limité.

L'approche fondamentale que l'on propose pour traiter ce problème repose sur les propriétés développées au paragraphe précédent. Nous aborderons successivement les formes accessibles, puis par dualité celles observables.

6-1 - Formes canoniques accessibles

Nous continuons à adopter les notations relatives aux systèmes discrets, toujours en raison de la signification précise des paramètres de Markov du système. Soit donc un système dynamique $\Sigma_n(\Phi, \Gamma, H)$ dont la paire (Φ, Γ) est accessible. Désignons par $n_i, i = 1, \dots, m$; et $\alpha_{ijk}, i = 1, \dots, m$; $j = 1, \dots, i$; $k = 1, \dots, n_i$ les invariants de (Φ, Γ) .

D'après la proposition III-3, il existe une matrice T régulière telle que :

$$T^{-1} \Phi T = \Phi^* \quad ; \quad \Gamma^* = T^{-1} \Gamma \text{ et } H^* = H T \quad (\text{III-62})$$

où :

$$\Phi_{ii}^* = \begin{bmatrix} 0 & & & \alpha_{i,i,1} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} ; \quad \Phi_{ij}^* = \begin{bmatrix} & & & \alpha_{i,j,1} \\ & & & \\ & & & \\ 0 & & & \alpha_{i,j,n_j} \end{bmatrix} \quad i < j ; \quad \Phi_{ij}^* = 0 \quad j > i \quad (\text{III-63})$$

et

$$\Gamma = [\gamma_1 \mid \dots \mid \gamma_m] ; \quad \gamma_i = e_{s_{i-1}} + 1 \quad (\text{III-64})$$

en notant e_k la $k^{\text{ième}}$ colonne de la matrice unité de dimension $n \times n$.

Déterminons maintenant H^* ; on sait que :

$$F_t = H \Phi^{t-1} \Gamma = H^* \Phi^{*t-1} \Gamma^* \quad (\text{III-65})$$

Partitionnons F_t de la manière suivante :

$$F_t = [\underline{f}_1^t \mid \dots \mid \underline{f}_m^t] \quad ; \quad \underline{f}_i = \begin{bmatrix} f_{li}^t \\ \vdots \\ f_{pi}^t \end{bmatrix} \quad (\text{III-66})$$

D'après la forme de Φ^* (III-63) et Γ^* (III-66) il vient, compte-tenu de (III-65) :

$$H_i^* = \begin{bmatrix} \underline{f}_i^1 & \dots & \underline{f}_i^{n_i} \end{bmatrix} \quad (\text{III-67})$$

où

$$H^* = \begin{bmatrix} H_1^* & \dots & H_i^* & \dots & H_m^* \\ \hline n_1 & & n_i & & n_m \end{bmatrix} \quad (\text{III-68})$$

On attachera à cette première structure canonique le nom de Luenberger en raison de la forme particulière de la paire (Φ^*, Γ^*) .

Remarque : Dans le cas particulier $m = p = 1$, on retrouve la forme canonique (3b) du théorème III-2.

Poursuivons en généralisant les résultats obtenus dans le cas mono-variable.

Définissons donc dans une transformation :

$$L' = \begin{bmatrix} L'_1 & & 0 \\ & \ddots & \\ 0 & & L'_m \end{bmatrix} \quad \text{où} \quad L'_i = \begin{bmatrix} 1 & a_1^i & \dots & a_{n_i-1}^i \\ & \ddots & & \\ 0 & & 1 & a_1^i \end{bmatrix} \quad (\text{III-69})$$

avec $a_k^i = -\alpha_{ii, n_i - k + 1}$

En utilisant les résultats du théorème III-2, il vient :

$$\Phi_{4b} = L'^{-1} \Phi^* L' ; \quad \Gamma_{4b} = L'^{-1} \Gamma^* ; \quad H_{4b} = H^* L' \quad (\text{III-70})$$

Les matrices Φ_{4b} , Γ_{4b} , H_{4b} ayant la structure suivante :

$$(\Phi_{4b})_{ii} = \begin{bmatrix} \alpha_{ii, n_i} & \dots & \alpha_{ii, 1} \\ 1 & & \\ & \ddots & \\ & & 1 & 0 \end{bmatrix} ; \quad (\Phi_{4b})_{ij} = L_i'^{-1} \Phi_{ij}^* L_j' = L_i'^{-1} \Phi_{ij} = \begin{bmatrix} \alpha'_{ij1} \\ \vdots \\ \alpha'_{ijn_j} \end{bmatrix} \quad \begin{matrix} i < j \\ i > j \end{matrix} \quad (\text{III-71})$$

$$\Gamma_{4b} = \Gamma^* \quad \dagger \quad (\text{III-72})$$

$$H_{4b} = [B_1 \vdots B_2 \vdots \dots \vdots B_m] ; \quad B_i = H_i^* L_i' \quad \dagger\dagger \quad (\text{III-73})$$

$\dagger$ Si certains n_i sont nuls la colonne de couplage est alors multipliée à gauche par $L_i'^{-1}$ au même titre que les colonnes de couplage dans Φ^* (Cf. (III-71))

$\dagger\dagger$ A ne pas confondre avec les B_i introduits au début du § 6 !

Les formes canoniques du type (1b) et (2b) (Cf. § 3) se déduisent de celles-ci, d'une manière triviale en permutant les états de chaque sous-espace X_i de dimension n_i .

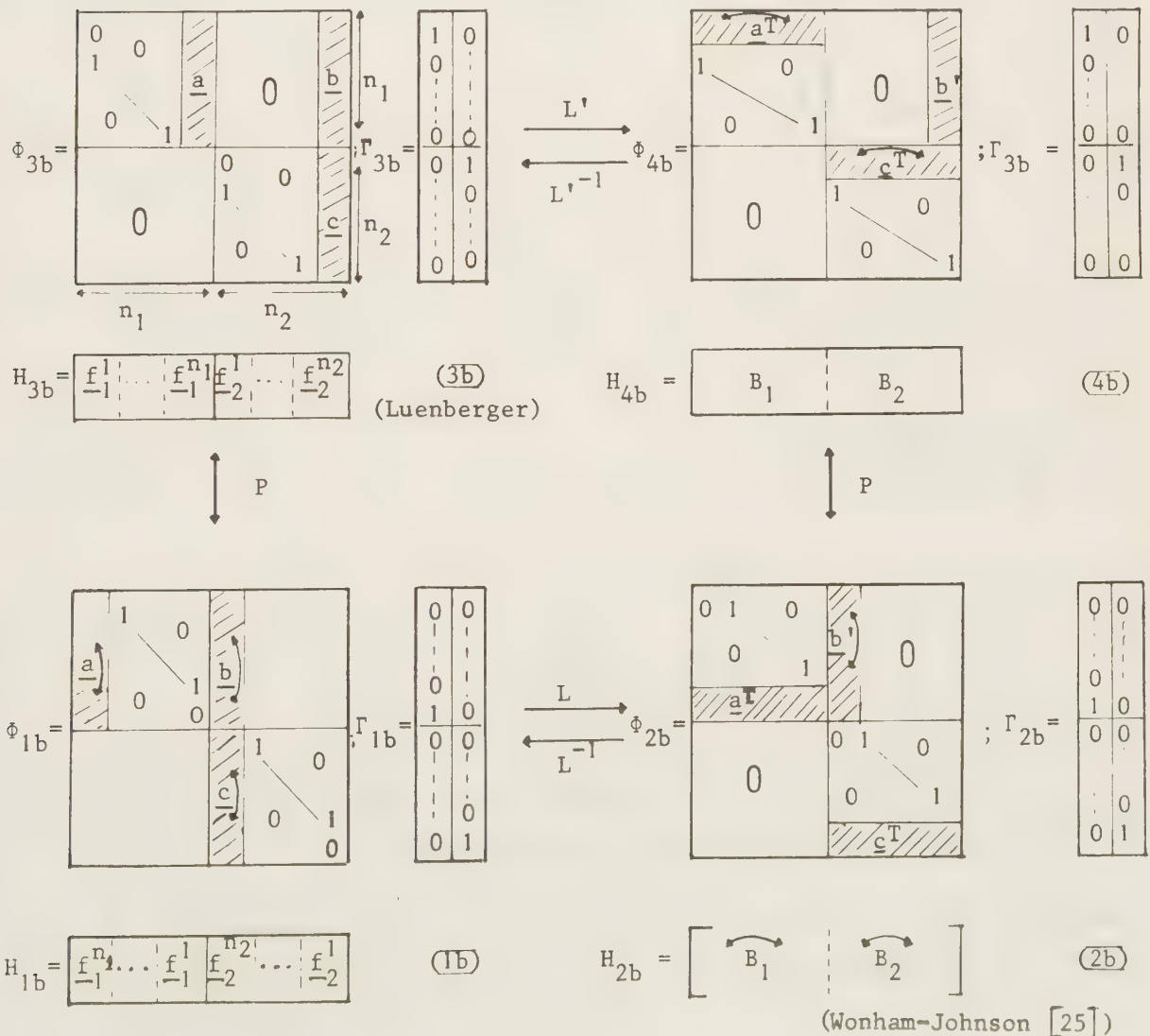
Les matrices de transformation P sont de la forme :

$$P = \begin{bmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_m \end{bmatrix} ; \quad P_i = \begin{bmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{bmatrix} \quad (\text{III-74})$$

Nous résumons ces résultats dans le théorème suivant :

Théorème III-7 : Les triplets $(\phi_{ib}, \Gamma_{ib}, H_{ib})$, $i = 1, \dots, 4$ définissent les formes équivalentes de la structure canonique accessible.

Ces quatre formes sont illustrées dans l'hypothèse $m = 2$, $n_1 > 0$, $n_2 > 0$



Avec les notations :

$$\underline{a} = \begin{pmatrix} \alpha_{111} \\ \vdots \\ \alpha_{11n_1} \end{pmatrix} ; \quad \underline{b} = \begin{pmatrix} \alpha_{121} \\ \vdots \\ \alpha_{12n_1} \end{pmatrix} ; \quad \underline{c} = \begin{pmatrix} \alpha_{221} \\ \vdots \\ \alpha_{22n_2} \end{pmatrix} ; \quad \underline{b}' = L_1'^{-1} \underline{b} ; \quad (\text{III-75})$$

... et symboles :

$$\curvearrowright \text{ permuter colonnes} ; \quad \updownarrow \text{ permuter lignes.}$$

6-2 - Formes canoniques observables

On admet donc, par hypothèse, que la paire (Φ, H) est maintenant observable. En raisonnant par dualité, la structure de la paire (Φ, H) est définie par un ensemble complet d'invariants indépendants n_i , $i = 1, \dots, p$; α_{ijk} , $j = 1, \dots, i$; $k = 1, \dots, n_i$; et il existe une transformation régulière T telle que :

$$T^{-1} \Phi T = \Phi^* ; \quad \Gamma^* = T^{-1} \Gamma \quad \text{et} \quad H^* = HT \quad (\text{III-76})$$

où

$$\Phi_{ii}^* = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & \searrow & 1 \\ & \alpha_{ii1} & \dots & \alpha_{iin_i} \end{bmatrix} ; \quad \Phi_{ij}^* = \begin{bmatrix} 0 \\ \hline \alpha_{ij1} \dots \alpha_{ijn_j} \end{bmatrix} \quad i > j ; \quad \Phi_{ij}^* = 0 \quad i < j \quad (\text{III-77})$$

$$H^* = \begin{bmatrix} \hline \underline{h}_i^T \hline \end{bmatrix} ; \quad \underline{h}_i = e_{s_{i+1}+1} \quad (\text{III-78})$$

$$\Gamma^* = \begin{bmatrix} \Gamma_1^* \uparrow n_1 \\ \vdots \\ \Gamma_p^* \uparrow n_p \end{bmatrix} ; \quad \Gamma_i^* = \begin{bmatrix} f_{i1}^1 & \dots & f_{im}^1 \\ \vdots & & \vdots \\ f_{i1}^{n_i} & \dots & f_{im}^{n_i} \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ \hline (f_i^t)^T \hline \vdots & \vdots \end{bmatrix} \updownarrow n_i \quad (\text{III-79})$$

Cette forme duale de Luenberger sera appelée forme canonique de Bucy en raison de la structure de la paire (Φ, H) .

Remarque : Dans le cas particulier où $m = p = 1$, on retrouve bien sûr la forme (2) du théorème III-1.

Il est clair qu'en effectuant la transformation :

$$L^{-1} = \begin{bmatrix} L^{-1} & 0 \\ 0 & L^{-1} \end{bmatrix} = P L'^{-1} P \quad (\text{III-80})$$

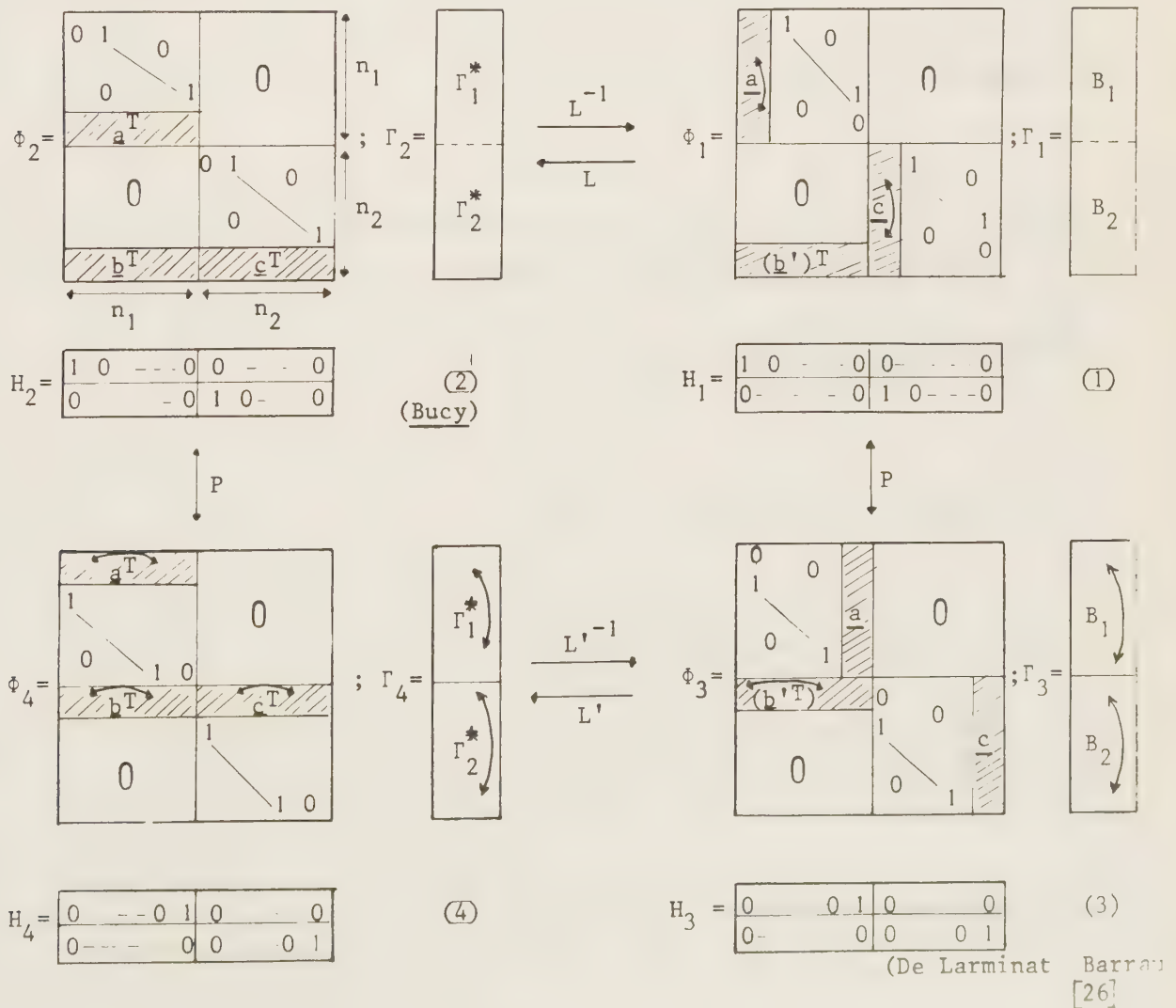
où L' et P sont définies respectivement en (III-69) et (III-74) en changeant m en p).

... on obtiendra la forme duale de (4b) soit la forme observable (1) en se référant aux notations du § 3. De même, les formes (3) et (4) se déduisent, par permutations, obtenues par la transformation P .

Ceci nous conduit à énoncer le ...

Théorème III-8 : Les triplets (Φ_i, Γ_i, H_i) ($i = 1, \dots, 4$) définissent les formes équivalentes de la structure canonique observable.

Ces quatre formes sont illustrées pour $p = 2, n_1 > 0, n_2 > 0$ et disposer de façon à faire correspondre aux mêmes emplacements "le dual" des formes accessibles.



Les notations et symboles sont ceux définis en (III-75).

Remarque : Ces deux derniers théorèmes s'appliquent, ipso-facto, aux systèmes dynamiques continus en changeant (Φ, Γ, H) en (A, B, C) .

7 - CONCLUSION

L'étude d'une structure particulière pour un opérateur linéaire, noté (A, B) , a permis de dégager une paramétrisation efficace des systèmes dynamiques, représentés par leur description interne. Cette paramétrisation est définie au moyen d'un ensemble complet d'invariants indépendants (n_i, α_{ijk}) dont le rôle est de caractériser la structure correspondante de $\Sigma, (n_i)$ et d'en préciser les éléments (α_{ijk}) .

Bien que reposant sur un raisonnement unique, (structure de l'opérateur (A, B)), les théorèmes (III-7) et (III-8), qui concrétisent deux développements duaux, généralisent complètement les résultats classiques mono-variables, (théorèmes (III-1) et (III-2)).

On a ainsi mis en évidence deux structures canoniques, l'une dans la forme accessible, l'autre dans la forme observable, qui exprimées dans les quatre "standards" relatifs aux matrices compagnons, englobent les résultats de Luenberger [14], Wonham et Johnson [25], Bucy [16] et De Larminat-Barraud [26].

BIBLIOGRAPHIE

- [1] F.R. GANTMACHER "Théorie des matrices" Tome I
Dunod - Paris 1966
- [2] R.W. BROCKETT "Finite dimensional linear systems"
John Wiley and Sons, Inc. New-York 1970
- [3] H.H. ROSENBROCK "State space and multivariable theory"
Nelson 1970
- [4] C.T. CHEN "Introduction to linear system theory"
Holt, Rinehart and Winston, Inc. 1970
- [5] A. GRAHAM "A note on a transformation between two
canonical forms in state space in terms
of the eigenvalues of the system matrix"
I.E.E.E. Trans. Aut. Cont. August 1968
p. 448
- [6] R.E. KALMAN "Mathematical Description of linear dynamical systems" - J. Soc. Ind. Appl. Math.
Ser. A On control, Vol. 1 n 2 (1963),
pp. 152-192
- [7] I.H. MUFTI "On the reduction of a system to canonical
(phase variable) form"
I.E.E.E. Trans. Aut. Cont. April 1965
pp 206-207
- [8] M.R. CHIDAMBARA "The transformation to (phase variable)
canonical form" - I.E.E.E. Trans. Aut.
Cont. - October 1965 pp. 492-495
- [9] L.M. SILVERMAN "Transformation of time variable systems
to canonical (phase variable) form"
I.E.E.E. Trans. Aut. Cont. - April 1966
pp 300-303
- [10] W.G. TUEL , Jr. "On the transformation to (phase variable)
canonical form" - I.E.E.E. Trans. Aut.
Cont. - July 1966 - p. 607
- [11] D.S. RANE "A simplified transformation to (phase
variable) canonical form". - I.E.E.E.
Trans. Aut. Cont. - July 1966 p. 608
- [12] C.D. JOHNSON "Another note on the transformation to
canonical (phase variable) form". I.E.E.E.
Trans. Aut. Cont. - July 1966 p. 609
- [13] B. RAMASWAMI "Transformation to the phase variable canonical
form" - I.E.E.E. Trans. Aut. Cont.
December 1968 - pp. 746-747

- [14] D.G. LUENBERGER "Canonical form for linear multivariable systems" - I.E.E.E. Trans. Aut. Cont. June 1967 - pp. 290-293
- [15] C.D. JOHNSON "A unified canonical for controllable and uncontrollable linear dynamical system" Int. J. Control 1971, Vol. 13 n 3 pp. 497-517
- [16] R.S. BUCY "Canonical forms for multivariable systems" I.E.E.E. Trans. Aut. Cont. - October 1968 pp. 567 - 569
- [17] C. BONIVENTO
R. GUIDORZI "Canonical input-output description of linear multivariable systems" - Ricerche di Automatica - Vol. 2 n 1 - January 1971.
- [18] S.J. ASSEO "Phase variable canonical transformation of multicontroller systems" - I.E.E.E. Trans. Aut. Cont. - February 1968 - pp. 129-131
- [19] Y.O. YUKSEL "Comments on - phase variable transformation of multicontroller systems -". I.E.E.E. Trans. Aut. Cont. - December 1970 - pp. 696-697
- [20] V.M. POPOV "Invariant description of linear, time invariant controllable systems" - Siam. J. Control - Vol. 10 n 8 - may 1972 - pp 252-264
- [21] J. RISSANEN "Bases of invariants and canonical forms for linear dynamical systems" - Automatica - Vol. 10 - pp. 175-182 - 1974.
- [22] W.R. WOLOVICH
P.L. FALB "On the structure of multivariable systems" Siam J. Control - Vol. 7 n 3 - August 1969
- [23] W.G. VOGT
R.W. GUNDERSON "Canonical realisation" Proc. 1st Asilomar Conf. on Circuits and Systems - pp. 783-790 - 1967
- [24] R.H. BARTEL
G.W. STEWART "Algorithm 432 : Solution of the matrix equation $AX + XB = C$ " C.A.C.M. - Vol. 15 Sept. 1972 - pp 820-823
- [25] C. MACHU "Formes canoniques pour les systèmes linéaires multivariables" - I.R.I.A. - Cahier n 3 - Novembre 1970
- [26] Ph. DE LARMINAT
A. BARRAUD "Un algorithme de réalisation minimale à partir de données d'entrées/sorties". CRAS t. 275 - 24 Juillet 1972 - série A pp. 315-318.

Chapitre IV

REALISATION CANONIQUES MINIMALES : ALGORITHMES



- 1 - Introduction
 - 2 - Réalisation d'une matrice de transfert rationnelle propre
 - 3 - Réalisation d'une suite de paramètres de Markov
 - 4 - Réalisation directe d'entrées/sorties quelconques (cas discret)
 - 5 - Réalisation directe d'entrées/sorties quelconques (cas continu)
 - 6 - Conclusion
- Bibliographie

Résumé :

L'objet de ce chapitre est de présenter un ensemble d'algorithmes de réalisations pour un système multivariable, réalisations ayant la structure canonique minimale.

L'idée générale autour de laquelle s'articulent ces différentes procédures repose uniquement sur le principe de sélection permettant d'élaborer la structure de l'opérateur (A,B) développée ch. III - § 5.

La simplicité et l'unicité de cette approche appliquée successivement à la réalisation d'un modèle externe (matrices de transfert, réponses impulsionnelles discrètes) puis directement d'entrées/sorties quelconques, définissent un point essentiel de notre contribution.

1 - INTRODUCTION

On vient de voir, chapitre III, que tout système dynamique décrit par une représentation interne pouvait être caractérisé par un ensemble d'invariants qui définissaient, d'une part, sa structure, et en représentaient, d'autre part, la meilleure paramétrisation vis à vis du problème de réalisation.

La détermination effective de ces paramètres est le thème que nous abordons ici. Nous distinguerons essentiellement deux problèmes suivant que l'on dispose ou non d'un modèle externe de ce système.

La réalisation d'une matrice de transfert rationnelle propre est développée, en premier, au § 2, à l'issue duquel on présente deux algorithmes "duaux" conduisant à deux formes canoniques minimales ayant respectivement la structure accessible ou observable.

Ces deux algorithmes sont, ensuite, modifiés, d'une manière appropriée (§ 3), pour traiter le cas où les données se présentent sous la forme d'une suite de paramètres de Markov.

Le problème, plus général et plus complexe, que représente la réalisation d'entrées/sorties quelconques, fait enfin l'objet des § 4 (systèmes discrets) et § 5 (systèmes continus). Outre cette distinction due à la nature des données, cette dernière approche nécessite, à la différence des points précédents, d'explicitier au préalable la relation d'entrée/sortie attachée à une structure canonique donnée § 4-1 et 4-2. De cette formulation même, découlent les algorithmes proprement dits (§ 4-4), moyennant certaines hypothèses sur les données : identifiabilité, réalisabilité (§ 4-3). Par ailleurs, en raison de la nature intrinséquement discrète des algorithmes (numériques) le traitement des systèmes continus ne pourra se faire que via des informations échantillonnées. Cette remarque a motivé la solution exposée, § 5, s'appuyant sur un modèle discret dont on déduit, alors, un certain modèle continu "équivalent" jouissant de la même structure canonique.

2 - REALISATION D'UNE MATRICE DE TRANSFERT RATIONNELLE PROPRE

Rappelons brièvement (Cf. Introduction) que c'est le premier problème abordé dans le domaine de la réalisation des systèmes [17], [18]. Depuis les premières méthodes basées sur la forme canonique de Smith des λ -matrices, on a pu enregistrer deux évolutions très nettes, l'une orientée vers l'élaboration d'algorithmes de calcul appropriés à une programmation relativement simple [19], l'autre visant à obtenir des réalisations sous une forme canonique. A notre sens, la technique développée par Chen et Mital [20] représente une approche intéressante dont s'inspire d'ailleurs l'algorithme I. La forme duale à laquelle conduit l'algorithme II représente un complément naturel aux résultats précédents.

D'une manière générale la différence essentielle avec [20] porte à la fois sur la démonstration adoptée ici et sur la mise en oeuvre numérique.

Soit donc une matrice rationnelle propre $F(z)$ de dimension $p \times m$, m désignant le nombre d'entrées, et p le nombre de sorties. Notons l'élément ij par :

$$F_{ij}(z) = \frac{N_{ij}(z)}{D_{ij}(z)} = \frac{b_0^{ij} z^{n_{ij}} + b_1^{ij} z^{n_{ij}-1} + \dots + b_{n_{ij}}^{ij}}{z^{n_{ij}} + a_1^{ij} z^{n_{ij}-1} + \dots + a_{n_{ij}}^{ij}} ; \quad \begin{cases} i = 1, \dots, p \\ j = 1, \dots, m \end{cases} \quad (IV-1)$$

de degré n_{ij} et supposé irréductible.

Par homogénéité avec les notations du Ch. III, on a implicitement adopté l'hypothèse des systèmes discrets, mais il est clair que tout ce qui suit s'applique, mutatis-mutandi, aux systèmes continus.

Rappelons (Cf. eq. III-8) que les paramètres de Markov (ici séquence de pondération) f_t^{ij} sont reliés à $F_{ij}(z)$, par la récurrence :

$$f_t^{ij} = \beta_t - \sum_{k=1}^{\min(t, n_{ij})} a_k^{ij} f_{t-k}^{ij} ; \quad t = 1, 2, \dots \quad (IV-2)$$

avec

$$f_0^{ij} = b_0^{ij} \quad \text{et} \quad \beta_t = \begin{cases} b_t^{ij} & \text{si } t \leq n \\ = 0 & \text{si } t > n \end{cases}$$

Nous développons maintenant deux algorithmes conduisant respectivement à une forme canonique minimale "observable" (forme de Bucy) et une forme canonique minimale "accessible" (forme de Luenberger 1) que nous noterons indifféremment $\Sigma(\Phi, \Gamma, H, D)$. Il est clair que, dans les deux cas, la transmission directe entrée-sortie n'est due qu'au terme b_o^{ij} , de sorte que :

$$D_{ij} = \begin{bmatrix} b_o^{ij} \\ 0 \end{bmatrix} \quad (IV-3)$$

C'est pourquoi on ne se bornera qu'à déterminer, dans ce qui suit, les matrices Φ , Γ et H .

Remarque : En ce qui concerne la dénomination des structures obtenues on se réfère au chapitre III § 6.

2-1. Algorithme 1 (forme observable)

Introduisons quelques notations. On désignera par v_j ($j = 1, \dots, m$) le degré du P.P.C.M. des polynômes $D_{ij}(z)$, $i = 1, \dots, p$.

Posons encore :

$$v = \sum_{j=1}^m v_j \quad \text{dimension du vecteur } \underline{f}_t^1 \text{ tel que}$$

$$(\underline{f}_t^1)^T = \left[\begin{array}{c|c|c|c} f_t^{i1}, \dots, f_{t+v_1-1}^{i1} & f_t^{i2}, \dots, f_{t+v_2-1}^{i2} & \dots & f_t^{im}, \dots, f_{t+v_m-1}^{im} \end{array} \right] \quad (IV-4)$$

... dont les composantes se déduisent des données via la récurrence (IV-2).

1 - Déterminer v_j , $j = 1, \dots, m$.

2 - Calculer $\underline{f}_1^1, \underline{f}_2^1, \dots$ jusqu'à ce qu'on rencontre le premier vecteur, linéairement dépendant des précédents, noté $\underline{f}_{-n_1+1}^1$. Ceci implique la relation :

$$\underline{f}_{-n_1+1}^1 = \sum_{k=1}^{n_1} \alpha_{1k} \underline{f}_k^1 \quad (IV-5).$$

3 - Calculer $\underline{f}_1^2, \underline{f}_2^2, \dots$ jusqu'à ce qu'on rencontre le premier vecteur linéairement dépendant des vecteurs $\underline{f}_1^1, \dots, \underline{f}_{-n_1}^1, \underline{f}_1^2, \underline{f}_2^2, \dots$ précédemment sélectionnés.

On notera ce vecteur $\underline{f}_{-n_2+1}^2$; il vient par conséquent :

$$\underline{f}_{-n_2+1}^2 = \sum_{k=1}^{n_1} \alpha_{21k} \underline{f}_k^1 + \sum_{k=1}^{n_2} \alpha_{22k} \underline{f}_k^2 \quad (IV-6)$$

Plus généralement $f_{n_i+1}^i$, $i = 1, \dots, p$ sera le premier vecteur linéairement dépendant des vecteurs déjà sélectionnés, tel que :

$$f_{n_i+1}^i = \sum_{j=1}^i \sum_{k=1}^{n_j} \alpha_{ijk} f_k^j \quad (\text{IV-7})$$

4 - En déduire la réalisation canonique minimale Φ , Γ , H définie par :

$$\Phi = \begin{bmatrix} \phi_{11} & & 0 \\ \phi_{21} & \phi_{22} & \\ \vdots & \vdots & \vdots \\ \phi_{p1} & \dots & \phi_{pp} \end{bmatrix}; \Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_p \end{bmatrix}; H = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \leftarrow \text{ligne } i$$

↑
colonne $s_{i-1} + 1$

(IV-8)

avec

$$\phi_{ii} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & \\ \alpha_{i1i} & \dots & \alpha_{in_i i} \end{bmatrix}; \phi_{ij} = \begin{bmatrix} 0 \\ \vdots \\ \alpha_{ij1} \dots \alpha_{ijn_j} \end{bmatrix}; \Gamma_i = \begin{bmatrix} f_{1i}^{i1} & \dots & f_{1i}^{im} \\ \vdots & & \vdots \\ f_{ni}^{i1} & \dots & f_{ni}^{im} \end{bmatrix};$$

$$\begin{cases} i = 1, \dots, p \\ j = 1, \dots, i \end{cases} \quad (\text{IV-9})$$

et

$$s_i = \sum_{l=1}^i n_l, \quad s_0 = 0 \quad (\text{IV-10})$$

Démonstration

Etant donnée la définition de v_j , $j = 1, \dots, m$ les triplets $(\phi_j^*, \Gamma_j^*, H_j^*)$ $j = 1, \dots, m$ sont des réalisations minimales des sous-systèmes mono-entrée (Cf. ch. III § 4 Théorème III-5 forme 3b) :

$$\phi_j^* = \begin{bmatrix} 0 & & x \\ 1 & & x \\ & \ddots & \vdots \\ & & 1 & x \end{bmatrix}; \Gamma_j^* = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; H_j^* = \begin{bmatrix} f_{1j}^{1j} & \dots & f_{1j}^{mj} \\ \vdots & & \vdots \\ f_{pj}^{1j} & \dots & f_{pj}^{mj} \end{bmatrix}$$
(IV-11)

où "x" dans ϕ_j^* désigne les coefficients du P.P.C.M. de $D_{ij}(z)$, $i = 1, \dots, p$.

On en déduit immédiatement une réalisation accessible, mais généralement non observable de $F(z)$ en effectuant la somme directe de ces réalisations partielles, soit :

$$\Phi^* = \begin{bmatrix} \phi_1^* & & 0 \\ & \ddots & \\ 0 & & \phi_m^* \end{bmatrix} ; \Gamma^* = \begin{bmatrix} \Gamma_1^* & & 0 \\ & \ddots & \\ 0 & & \Gamma_m^* \end{bmatrix} ; H^* = H_1^*, \dots, H_m^* \quad (IV-12)$$

On reconnaîtra que les lignes de H^* ne sont autres que les vecteurs $(\underline{f}_1^i)^T$; $i = 1, \dots, p$. Plus généralement, compte-tenu de (IV-11), les lignes de $H^*(\Phi^*)^t$ sont les vecteurs $(\underline{f}_{t+1}^i)^T$; $i = 1, \dots, p$.

Par suite, les étapes 2 et 3 de l'algorithme 1 consistent à déterminer la structure de l'opérateur (Φ^*, H^*) au sens défini au chapitre III - § 5. Combinant ce résultat et celui du théorème II-9, (raisonnement dual de celui tenu au § 5-3 Ch. III), les paramètres n_i et α_{ijk} sont les invariants de la partie observable de Σ^* , puisque les vecteurs $\underline{f}_t^i$ sélectionnés sont ceux qui engendrent l'espace observable.

En conséquence (Φ, Γ, H) est une réalisation canonique de $F(z)$ minimale puisqu'étant la partie observable de (Φ^*, Γ^*, H^*) , qui est, elle-même, une réalisation accessible.

Remarque : Si un n_i est nul, la chaîne $\underline{f}_t^i$ disparaît ainsi que les éléments ϕ_{ij} , $j = 1, \dots, i$ de Φ et Γ_i de Γ . Les couplages qui devaient être contenus dans ϕ_{ij} , $j = 1, \dots, i - 1$ sont reportés dans les s_{i-1} premières positions de la ligne i de H .

Exemple : $n_1 = 3$; $n_2 = 0$; $n_3 = 2$ ($m = 2$)

$$\Phi = \begin{bmatrix} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & 0 \\ \alpha_{111} & \alpha_{112} & \alpha_{113} & & \\ 0 & 0 & 0 & 0 & 1 \\ \alpha_{311} & \alpha_{312} & \alpha_{313} & \alpha_{331} & \alpha_{332} \end{bmatrix} ; \Gamma = \begin{bmatrix} f_1^{11} & f_1^{12} \\ f_2^{11} & f_2^{12} \\ f_3^{11} & f_3^{12} \\ f_1^{31} & f_1^{32} \\ f_2^{31} & f_2^{32} \end{bmatrix} ; H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_{211} & \alpha_{212} & \alpha_{213} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Mise en oeuvre : L'algorithme repose, essentiellement, sur la détermination de la dépendance linéaire ou non, d'une suite de vecteurs. C'est donc ce point que nous examinons en premier.

Soit $\{\underline{x}_k\}$ une suite de vecteurs dont on veut déterminer l'indice du premier vecteur linéairement dépendant des vecteurs précédents et, éventuellement, les coefficients qui expriment cette dépendance. Pour y parvenir, nous allons utiliser les transformations orthogonales élémentaires de Householder de la forme :

$$Q = I - 2 \underline{u} \underline{u}^T \quad ; \quad ||\underline{u}|| = 1 \quad (\text{IV-13})$$

Supposons $||\underline{x}_1|| \neq 0$ et soit Q_1 la transformation définie par la relation[†] :

$$\underline{x}_1^* = Q_1 \underline{x}_1 \triangleq \begin{bmatrix} a_1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{IV-14})$$

Posons maintenant :

$$\underline{x}_2' = Q_1 \underline{x}_2 = \begin{bmatrix} a_2 \\ \vdots \\ b_2 \end{bmatrix} \quad (\text{IV-15})$$

Si $||\underline{b}_2|| = 0$ (ou $||\underline{b}_2|| \ll |a_2|$) alors $\underline{x}_2$ dépend linéairement de $\underline{x}_1$ et

$$\underline{x}_2 = \frac{a_2}{a_1} \underline{x}_1.$$

Dans le cas contraire on appliquera à $\underline{x}_2'$ la transformation Q_2 telle que :

$$\underline{x}_2^* = Q_2 \underline{x}_2' = Q_2 Q_1 \underline{x}_2 = \begin{bmatrix} x \\ x \\ 0 \\ 0 \end{bmatrix} \quad (\text{IV-16})$$

... de sorte que les deux vecteurs $\underline{x}_1$, $\underline{x}_2$ ont été transformé de façon à former le début d'une matrice triangulaire supérieure :

$$\begin{bmatrix} \underline{x}_1 & \vdots & \underline{x}_2 \end{bmatrix} \xrightarrow{Q_2 Q_1} \begin{bmatrix} x & x \\ 0 & x \\ \vdots & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \underline{x}_1^* & \underline{x}_2^* \end{bmatrix} \quad (\text{IV-17})$$

dont les éléments diagonaux sont nécessairement non nuls.

[†] On renvoie le lecteur à [1] pour l'existence de Q_1 et la détermination correspondante de $\underline{u}$.

Plus généralement supposons que les p premiers vecteurs aient été trouvés linéairement indépendants : on a donc pu construire p transformations $Q_1, Q_2, \dots, Q_p$ tels qu'ils ont été transformés successivement en $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_p^*$ de façon à former la matrice (IV-18).

$$Q_p \dots Q_1 \cdot \begin{bmatrix} \vdots \\ \underline{x}_1 \\ \vdots \\ \underline{x}_p \\ \vdots \end{bmatrix} = \begin{bmatrix} \text{matrice triangulaire} \\ 0 \end{bmatrix} \begin{bmatrix} \vdots \\ \underline{x}_1^* \\ \vdots \\ \underline{x}_p^* \\ \vdots \end{bmatrix} \quad (\text{IV-18})$$

Appliquons maintenant à $\underline{x}_{p+1}$ les transformations $Q_1, Q_2, \dots, Q_p$:

$$\underline{x}_{p+1}' \triangleq Q_p Q_{p-1} \dots Q_1 \underline{x}_{p+1} = \begin{bmatrix} \underline{a}_{p+1} \\ \vdots \\ \underline{b}_{p+1} \end{bmatrix} \quad (\text{IV-19})$$

Il est clair que si $||\underline{b}_{p+1}|| = 0$ (ou $\ll ||\underline{a}_{p+1}||$) $\underline{x}_{p+1}$ est linéairement dépendant de $\underline{x}_1, \dots, \underline{x}_p$ et les coefficients $\underline{\alpha}$ exprimant cette dépendance sont solution du système triangulaire extrait de (IV-18).

$$\begin{bmatrix} \text{matrice triangulaire} \\ 0 \end{bmatrix} \underline{\alpha} = \begin{bmatrix} \underline{a}_{p+1} \end{bmatrix} \quad (\text{IV-20})$$

Dans le cas contraire on détermine Q_{p+1} telle que :

$$\underline{x}_{p+1}^* = Q_{p+1} \underline{x}_{p+1}' \triangleq \begin{bmatrix} \underline{x} \\ \vdots \\ \underline{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{IV-21})$$

Le processus est, bien évidemment, limité à p égal à la dimension de $\underline{x}$.

Remarques :

- Cette technique est bien connue pour sa stabilité numérique, et présente l'avantage de donner une interprétation précise de l'approximation que l'on est toujours amené à faire. En effet, sauf hasard, $||\underline{b}||$ sera différent de zéro. La solution $\underline{\alpha}$ donné par (IV-20) correspond à l'approximation, au sens des moindres carrés, de $\underline{x}_{p+1}$ à l'aide de $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$. De plus $||\underline{b}||$ représente ici la distance de $\underline{x}_{p+1}$ à l'espace engendré par $\underline{x}_1, \dots, \underline{x}_p$, de sorte qu'il est facile de conclure quant au test d'arrêt.

- Du point de vue pratique, les matrices Q ne sont jamais calculées explicitement mais seuls les vecteurs $\underline{u}$ correspondants. Ceux-ci sont mémorisés à un élément près à la place des zéros qu'ils font apparaitre dans $\underline{x}$; Cf. [1].

Le deuxième point à développer concerne la détermination des v_j , $j = 1, \dots, m$. Si les pôles des systèmes sont connus, le problème est trivial. Dans le cas contraire v_j peut être déterminé comme étant le nombre de vecteurs $\underline{g}_t^j$ ($t = 1, 2, \dots, v_j$) linéairement indépendants où :

$$(\underline{g}_t^j)^T = \left[f_t^{1j}, \dots, f_{t+n_{1j}-1}^{1j} \mid f_t^{2j}, \dots, f_{t+n_{2j}-1}^{2j} \mid \dots \mid f_t^{pj}, \dots, f_{t+n_{pj}-1}^{pj} \right] \quad (IV-22)$$

On sait (Cf. Ch. III - § 3 Théorème III-1 forme (2)) en effet qu'une réalisation minimale de $F_{ij}(z)$ est donnée par :

$$\tilde{\Phi}_j = \begin{bmatrix} 0 & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ -a_{n_{ij}}^{ij} & & & & -a_1^{ij} \end{bmatrix}; \quad \tilde{\Upsilon}_{ij} = \begin{bmatrix} f_1^{ij} \\ \vdots \\ f_{n_{ij}}^{ij} \end{bmatrix}; \quad \tilde{H}_{ij} = [1 \quad 0 \quad \dots \quad 0] \quad (IV-23)$$

On en déduit alors une réalisation observable mono-entrée, par somme directe, soit :

$$\tilde{\Phi}_j = \begin{bmatrix} \tilde{\Phi}_{1j} & 0 \\ 0 & \tilde{\Phi}_{pj} \end{bmatrix}; \quad \tilde{\Upsilon}_j = \begin{bmatrix} \tilde{\Upsilon}_{1j} \\ \vdots \\ \tilde{\Upsilon}_{pj} \end{bmatrix}; \quad \tilde{H}_j = \begin{bmatrix} \tilde{H}_{1j} & 0 \\ 0 & \tilde{H}_{pj} \end{bmatrix} \quad (IV-24)$$

... dont les colonnes de la matrice d'accessibilité sont précisément $\underline{g}_t^j$. En appliquant, une fois de plus, les théorèmes de structure et de décomposition canonique, le nombre de vecteurs $\underline{g}_t^j$ linéairement indépendants correspond à la dimension de la réalisation minimale de chaque sous-système mono-entrée, c'est à dire au degré du P.P.C.M. des D_{ij} , $i = 1, \dots, p$, c'est à dire v_j .

Remarque : La connaissance des v_j n'est pas fondamentale pour l'algorithme 1 puisque les invariants du système pourront, tout aussi bien, être déterminés à partir de vecteurs $\underline{f}_t^i$ dans lesquels on aura substitué v_j par sa borne supérieure $v_j' = \sum_{i=1}^p n_{ij}$. La raison en est simple. En opérant ainsi on ajoute à la matrice dont les colonnes sont les vecteurs $\underline{f}_t^i$, des lignes linéairement dépendantes ce qui ne peut en changer la structure.

L'économie de calcul que l'on fait en prenant v_j' se paie au niveau de la détermination des invariants du système. Le choix d'un compromis dépend de chaque cas, et en particulier des degrés n_{ij} .

2-2. Algorithme 2 (forme accessible)

Cet algorithme est le dual du précédent, en ce sens qu'il conduira à la forme canonique duale de celle de Bucy, en l'occurrence une forme de Luenberger. A la différence du premier on sera amené à raisonner sortie par sortie et non plus entrée par entrée.

Soit donc μ_i , $i = 1, \dots, p$ le degré du P.P.C.M. des polynômes $D_{ij}(z)$, $j = 1, \dots, m$. On notera

$$\mu = \sum_{i=1}^p \mu_i$$

Définissons encore un vecteur $\underline{f}_t^j$ de dimension μ par :

$$(\underline{f}_t^j)^T = \left[f_t^{1j}, \dots, f_{t+\mu_1-1}^{1j} \mid f_t^{2j}, \dots, f_{t+\mu_2-1}^{2j} \mid \dots \mid f_t^{pj}, \dots, f_{t+\mu_p-1}^{pj} \right] \quad (\text{IV-25})$$

1 - Déterminer μ_i , $i = 1, \dots, p$

2 - Déterminer n_1 le nombre de vecteurs successifs $\underline{f}_1^1, \underline{f}_2^1, \dots, \underline{f}_{n_1}^1$ linéairement indépendants.

Soit donc :

$$\underline{f}_{n_1+1}^1 = \sum_{k=1}^{n_1} \beta_{11k} \underline{f}_k^1 \quad (\text{IV-26})$$

3 - Plus généralement pour $j = 2, \dots, m$ déterminer le nombre n_j de vecteurs successifs $\underline{f}_t^j$ linéairement indépendants de ceux déjà sélectionnés ; ce qui conduit à la relation :

$$\underline{f}_{n_j+1}^j = \sum_{i=1}^j \sum_{k=1}^{n_i} \beta_{ijk} \underline{f}_k^i \quad (\text{IV-27})$$

4 - En déduire la réalisation canonique minimale Φ , Γ , H définie par :

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \dots & \Phi_{1m} \\ & \Phi_{22} & & \\ & & \ddots & \\ 0 & & & \Phi_{mm} \end{bmatrix} ; \Gamma = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \leftarrow \begin{array}{l} \text{ligne} \\ s_{j-1}+1 \end{array} \quad ; H = \begin{bmatrix} H_1 & H_2 & \dots & H_m \end{bmatrix} \quad (\text{IV-28})$$

↑
colonne j

avec

$$\Phi_{ii} = \begin{bmatrix} 0 & & \beta_{ii1} \\ 1 & & \vdots \\ & 1 & \beta_{iin_i} \end{bmatrix} ; \Phi_{ij} = \begin{bmatrix} \beta_{ij1} \\ \vdots \\ \beta_{ijn_i} \end{bmatrix} ; H_j = \begin{bmatrix} f_1^{1j} & \dots & f_{n_j}^{1j} \\ \vdots & & \vdots \\ f_1^{pj} & \dots & f_{n_j}^{pj} \end{bmatrix} \quad (\text{IV-29})$$

$j = 1, \dots, m$
 $i = 1, \dots, j$

et :

$$s_j = \sum_{l=1}^j n_l ; s_0 = 0 \quad (IV-30)$$

Démonstration

Etant donné la définition des μ_i , $i = 1, \dots, p$ les triplets $(\phi_i^*, \Gamma_i^*, H_i^*)$ sont des réalisations minimales des sous-systèmes mono-sortie.

$$\phi_i^* = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ x & x & \dots & x \end{bmatrix} ; \Gamma_i^* = \begin{bmatrix} f_{11}^{i1} & \dots & f_{1\mu_i}^{im} \\ \vdots & & \vdots \\ f_{\mu_i 1}^{i1} & \dots & f_{\mu_i \mu_i}^{im} \end{bmatrix} ; H_i^* = [1 \ 0 \ \dots \ 0] \quad (IV-31)$$

... dont la somme directe constitue une réalisation observable de $F(z)$, soit :

$$\phi^* = \begin{bmatrix} \phi_1^* & & 0 \\ & \ddots & \\ 0 & & \phi_p^* \end{bmatrix} ; \Gamma^* = \begin{bmatrix} \Gamma_1^* \\ \vdots \\ \Gamma_p^* \end{bmatrix} ; H^* = \begin{bmatrix} H_1^* & & 0 \\ & \ddots & \\ 0 & & H_p^* \end{bmatrix} \quad (IV-32)$$

La encore, on vérifie facilement que les colonnes de la matrice d'accessibilité ne sont autres que les vecteurs $\underline{f}_t^j$. Par suite les étapes 2 et 3 consistent à déterminer la structure invariante de $F(z)$. En combinant ce résultat et celui du théorème II-3 (se rappeler le raisonnement Ch. III § 5-3), le triplet (ϕ, Γ, H) constitue alors la forme canonique de la partie accessible de la réalisation (ϕ^*, Γ^*, H^*) observable, et par conséquent est une réalisation minimale.

Remarques : - Le cas particulier d'un $n_i = 0$ se traite d'une manière analogue à l'algorithme 1.

- En ce qui concerne sa mise en oeuvre, elle se fait suivant la même technique. La détermination des μ_i peut être conduite en déterminant le nombre de vecteurs $\underline{g}_t^i$ linéairement indépendants défini par :

$$(\underline{g}_t^i)^T = \left[f_t^{i1}, \dots, f_{t+n_{i1}-1}^{i1} \mid f_t^{i2}, \dots, f_{t+n_{i2}-1}^{i2} \mid \dots \mid f_t^{im}, \dots, f_{t+n_{im}-1}^{im} \right] \quad (IV-33)$$

raisonnement basé, cette fois, sur la construction d'une réalisation minimale des sous-systèmes mono-sortie à partir de la somme directe des réalisations minimales "accessibles" des systèmes élémentaires $F_{ij}(z)$, $j = 1, \dots, m$.

- Le compromis entre cette approche et le choix par défaut de la borne supérieure $\mu_1^! = \sum_{j=1}^m n_{ij}$ demeure un problème d'espèce.

- Le choix entre l'algorithme 1 et 2 dépend du besoin de disposer d'une forme du type "observable" ou "accessible". Ces deux structures (IV-8) et (IV-28) ont été retenues parce que particulièrement simples à déterminer ; mais il est clair que les autres formes sont triviales à obtenir à partir de celles-ci (permutation ou/et résolution des systèmes triangulaires $L_i, n_i \times n_i$).

3 - REALISATION D'UNE SUITE DE PARAMETRES DE MARKOV

Dans ce domaine, l'algorithme de Ho [2] est longtemps resté la procédure "classique". Néanmoins l'absence de structure particulière des réalisations correspondantes représente l'inconvénient majeur de cette technique, par ailleurs très proches de la notion théorique de factorisation (Cf. Ch. II - § 6).

Nous avons suivi ici une autre démarche d'esprit analogue à la méthode de Bucy [21], mais là encore différente quant à la démonstration du résultat (forme observable minimale) et à la mise en oeuvre de la procédure (Algorithme 3). La forme duale (structure accessible) sera par ailleurs obtenue directement par l'Algorithme 4.

Cependant à la différence du problème précédent, où la réalisabilité du modèle externe était à priori assurée, en raison même de la forme des données (transmittances rationnelles), il convient, ici, d'apporter quelques précisions. Rappelons à ce propos deux résultats préliminaires classiques [2], [3].

Lemme IV-1 : La dimension d'une réalisation irréductible de la suite infinie $\{F_k\}$ est égale au rang de la matrice de Hankel généralisée $\mathcal{H}_\infty$, où :

$$\mathcal{H}_r = \begin{bmatrix} F_1 & F_2 & \dots & F_r \\ F_2 & F_3 & & \vdots \\ \vdots & & & \vdots \\ F_r & \dots & & F_{2r-1} \end{bmatrix} \quad (\text{IV-34})$$

Lemme IV-2 : La matrice $\mathcal{H}_\infty$ est de rang fini, si, et seulement si, il existe un nombre fini p de coefficients (non tous nuls) α_i , tels que :

$$F_k = \sum_{i=1}^p \alpha_i F_{k-i} \quad k = p+1, \dots \quad (\text{IV-35})$$

et $\rho[\mathcal{H}_t] = \rho[\mathcal{H}_p] = n \quad \forall t \geq p$

Supposons la suite $\{F_k\}$ induite par une transmittance rationnelle $F(z)$ dont le degré du P.P.C.M. de tous les dénominateurs soit r , alors $\rho[\mathcal{L}_r] = n$ [2]. Dans ce cas il suffit d'un nombre fini de paramètres de Markov pour construire une réalisation minimale de la suite infinie.

En l'absence d'une telle information, et ne connaissant qu'une sous-suite finie $\{F_1, \dots, F_N\}$ on ne pourra parler que de réalisation partielle, c'est à dire ne portant que sur l'identité d'un nombre fini de paramètres de Markov. On appellera à cet effet le :

Théorème IV-1 [3] : Etant donnée la suite finie $\{F_1, F_2, \dots, F_{2r+1}\}$, telle que :

$$\rho[\mathcal{L}_r] = \rho[\mathcal{L}_{r+1}] = n \quad (\text{IV-36})$$

alors il correspond à cette séquence une réalisation partielle minimale de dimension n dont les $(2r+1)$ premiers paramètres de Markov sont $\{F_k\}$ - $k = 1, \dots, 2r+1$.

On admettra dans la suite que les conditions de ce théorème sont remplies, c'est à dire en pratique que le système admet une réalisation finie et que la suite de paramètres de Markov, dont on dispose, est suffisamment longue.

On est maintenant à même de définir deux algorithmes directement dérivés de ceux exposés précédemment (§ 2).

3-1. Algorithme 3 (Forme observable)

1 - Etant donné $F_1, F_2, \dots, F_N$; poser $r = \left\lfloor \frac{N-1}{2} \right\rfloor$; ($[x]$ partie entière de x).

2 - Appliquer les étapes 2 et 3 de l'algorithme 1 aux vecteurs $\underline{f}_t^i$, $i = 1, \dots, p$ $t \leq r$, avec ici :

$$(\underline{f}_t^i)^T = \left[f_t^{i1}, \dots, f_t^{im} \mid f_{t+1}^{i1}, \dots, f_{t+1}^{im} \mid \dots \mid f_{t+r-1}^{i1}, \dots, f_{t+r-1}^{im} \right] \quad (\text{IV-37})$$

sachant que :

$$F_t = \begin{bmatrix} f_t^{11} & \dots & f_t^{1m} \\ \vdots & & \vdots \\ f_t^{p1} & \dots & f_t^{pm} \end{bmatrix} \quad (\text{IV-38})$$

Remarque : Ceci revient à poser $v_1 = v_2 = \dots = v_m = r$ dans l'algorithme 1, et à réordonner les composantes de $\underline{f}_t^i$ définies alors en (IV-4).

3 - En déduire la construction, suivant l'étape 4 de l'algorithme 1, d'une réalisation canonique minimale (partielle) Φ, Γ, H dans la forme standard observable de Bucy (IV-8).

Démonstration :

Par hypothèse la suite $\{F_t\}$ a une réalisation finie et l'on dispose, à priori, d'un nombre N suffisant de paramètres pour construire une réalisation de cette séquence.

Ceci signifie donc que :

$$\rho [\mathcal{H}_r] = \rho [\mathcal{H}_{r+1}]$$

... et donc qu'il existe des quantités α_i telles que :

$$F_t = \sum_{i=1}^r \alpha_i F_{t-i} \quad t = r+1, \dots, 2r+1 \quad (\text{IV-39})$$

Ces paramètres F_t peuvent donc être induits par une matrice rationnelle dont le degré du P.P.C.M. de tous les dénominateurs est r . A cette matrice correspond des degrés $v_1, \dots, v_m \leq r$. Par conséquent travailler sur les vecteurs $\underline{f}_t^i$ définis en (IV-4) ou en (IV-37) ne change strictement rien. Dans le dernier cas ne connaissant pas explicitement les v_j ; $j = 1, \dots, m$ on introduit, dans la matrice constituée des lignes $(\underline{f}_t^i)^T$, des colonnes linéairement dépendantes en nombre $\sum_{j=1}^n (v_j - r)$; ce qui ne peut modifier la structure du système, et donc nous ramène au cas précédent.

3-2. Algorithme 4 (forme accessible)

1 - Etant donné $F_1, F_2, \dots, F_N$; poser $r = \left\lceil \frac{N-1}{2} \right\rceil$

2 - Appliquer les étapes 2 et 3 de l'algorithme 2 aux vecteurs $\underline{f}_t^j$, $j = 1, \dots, m$, $t < r$, avec :

$$(\underline{f}_t^j)^T = \begin{bmatrix} 1j & & pj & 1j & & pj & & 1j & & pj \\ f_t & \dots & f_t & f_{t+1} & \dots & f_{t+1} & \dots & f_{t+r-1} & \dots & f_{t+r-1} \end{bmatrix} \quad (\text{IV-40})$$

Remarque : Ceci revient à poser $\mu_1 = \mu_2 = \dots = \mu_p = r$ et à réordonner les composantes du vecteur défini en (IV-25).

3 - En déduire, suivant l'étape 4 de l'algorithme 2, la construction d'une réalisation canonique minimale (partielle) Φ , Γ , H dans la forme standard accessible de Luenberger (IV-28).

Démonstration : Elle suit le même raisonnement dans le contexte dual du précédent.

Mise en oeuvre :

Elle est plus simple que pour les algorithmes 1 et 2 puisqu'il n'y a plus à calculer les paramètres F_t ni à déterminer les quantités v_j ou μ_i . Elle ne fait donc appel qu'à la procédure décrite pour déterminer soit l'indépendance linéaire des vecteurs soit leurs coefficients de dépendance.

4 - REALISATION DIRECTE A PARTIR D'ENTREES/SORTIES QUELCONQUES (Cas discret)

Cette dernière approche est fondamentalement différente des précédentes, par les difficultés nouvelles qu'elle introduit en raison même de la nature des données. Systèmes discrets et continus doivent être considérés ici comme deux entités, cette dichotomie reposant sur l'impossibilité de traiter complètement, avec des moyens numériques, des informations analogiques (continues). La difficulté et le coût que représente une simulation précise d'un système continu explique et justifie la voie (plus ou moins artificielle) qui consiste à modéliser des informations échantillonnées au moyen d'un système discret, puis, d'en déduire si besoin est, un modèle continu "équivalent" (Cf § 5).

En s'en tenant aux systèmes discrets, la construction d'une réalisation minimale d'une séquence d'entrées/sorties $\{\underline{u}_t, \underline{y}_t\}$, $t = t_0, \dots, T$, est néanmoins un problème dont les premières solutions sont encore récente [22], [23]. Les deux premiers algorithmes conduisant à une des formes canonique, développés Ch. III, sont dûs (indépendamment et à peu près à la même époque) à Ackermann [24] (forme (2)) et à Barraud et De Larminat [5], [10], (forme (3)).

Là encore, notre effort a porté sur l'unification des démonstrations et de la mise en oeuvre d'algorithmes permettant d'obtenir un modèle dans l'une des quatre formes canoniques "observables" d'un système $\Sigma(\Phi, \Gamma, H, D)$. Il convient de noter, ici, qu'on ne peut plus construire, directement à partir d'entrées/sorties, des réalisations ayant la structure "accessible". En effet, sur le plan algorithme, on a vu précédemment (§ 2 et 3) que les structures accessibles ou observables étaient respectivement obtenues en décomposant implicitement le système multivariable original en m sous-systèmes mono-entrée,

ou en p sous-systèmes mono-sortie. Or il est clair qu'on ne peut maintenant décomposer le système original qu'en sous-systèmes mono-sortie, dans la mesure où à priori toutes les entrées agissent sur une sortie donnée. Par ailleurs, si l'on se rappelle la décomposition canonique d'un système (Cf. Ch. II - § 6) la figure 6 page 45 illustre clairement le fait que la modélisation étant faite à partir des mesures entrées-sorties, ces dernières pourront contenir la contribution d'un état observable mais non accessible.

Avant d'aborder l'élaboration même de ces algorithmes, il convient de définir les relations d'entrées/sorties associées aux structures canoniques observables d'un système Σ , préliminaire nécessaire à tout autre développement.

Nous considérerons ainsi successivement les formes observables (2) et (1) ainsi que leur version "permutée" notée respectivement (4) et (3).

4-1. Relations entrée/sortie associées à la forme canonique observable (2)-(4)

Rappelons que la réalisation canonique (2) présente la structure suivante :

$$\left\{ \begin{array}{l} \underline{x}_{t+1} = \begin{bmatrix} \Phi_{11} & & & 0 \\ \Phi_{21} & \Phi_{22} & & \\ \vdots & & \ddots & \\ \Phi_{p1} & \Phi_{p2} & \dots & \Phi_{pp} \end{bmatrix} \underline{x}_t + \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_p \end{bmatrix} \underline{u}_t \\ \\ \underline{y}_t = \begin{bmatrix} \vdots \\ 0 \dots 0 \ 1 \ 0 \dots 0 \\ \vdots \end{bmatrix} \underline{x}_t + D \underline{u}_t \end{array} \right. \quad \text{(IV-42)}$$

$\uparrow$ $i^{\text{ème}}$ ligne $\uparrow$ $s_{i-1} + 1^{\text{ème}}$ colonne $\uparrow$

où les blocs Φ_{ij} de dimension $n_i \times n_j$ sont donnés par :

$$\Phi_{ij} = \begin{bmatrix} 0 \\ \vdots \\ \alpha_{ij1} \dots \alpha_{ijn_j} \end{bmatrix} \quad i > j \quad \text{et} \quad \Phi_{ii} = \begin{bmatrix} 0 \ 1 \dots 0 \\ 0 \dots \ddots \ 1 \\ \alpha_{ii1} \dots \alpha_{iin_i} \end{bmatrix} \quad \text{(IV-43)}$$

+ On rappelle que $s_i = \sum_{j=1}^i n_j$, $s_0 = 0$

Nous allons d'abord considérer l'hypothèse où $m = p = 1$ pour illustrer plus clairement le mécanisme d'élimination des états que nous exposons maintenant et qui sera ensuite généralisé au cas multivariable, équation (IV-12).

$$\begin{cases} \underline{x}_{t+1} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & \\ \alpha_1 & \dots & \dots & \alpha_n \end{bmatrix} \underline{x}_t + \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} u_t \\ y_t = [1 \ 0 \ \dots \ 0] \underline{x}_t + d u_t \end{cases} \quad (IV-44)$$

Remarquons que :

$$\Phi \underline{x}_t = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & \\ \alpha_1 & \dots & \dots & \alpha_n \end{bmatrix} \cdot \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^n \\ x_t^n \end{bmatrix} = \begin{bmatrix} x_t^2 \\ \vdots \\ x_t^n \\ \xi \end{bmatrix} \quad (IV-45)$$

où

$$\xi = \sum_{i=1}^n \alpha_i x_t^i \triangleq \underline{\alpha}^T \underline{x}_t \quad (IV-46)$$

Calculons maintenant la sortie aux instants $t, t+1, t+2, \dots, t+n-1$. Il vient :

$$\begin{cases} y_t = (10 \ \dots \ 0) \underline{x}_t + d u_t \\ y_t = x_t^1 + d u_t \end{cases} \quad (a) \quad (IV-47)$$

$$\begin{cases} y_{t+1} = (10 \ \dots \ 0) (\Phi \underline{x}_t + \Gamma u_t) + d u_{t+1} \\ y_{t+1} = (10 \ \dots \ 0) \left[\begin{pmatrix} x_t^2 \\ \vdots \\ x_t^n \\ \xi \end{pmatrix}_t + \Gamma u_t \right] + d u_{t+1} \\ y_{t+1} = x_t^2 + \gamma_1 u_t + d u_{t+1} \end{cases} \quad (b)$$

d'où, première relation :

$$y_{t+n} = \underline{\alpha}^T \underline{x}_t + \gamma_n u_t + \gamma_{n-1} u_{t+1} + \dots + \gamma_1 u_{t+n-1} + d u_{t+n} \quad (\text{IV-49})$$

... qui se transforme après élimination de $\underline{x}_t$ (IV-48) :

$$y_{t+n} = \alpha_1 y_t + \alpha_2 y_{t+1} + \dots + \alpha_n y_{t+n-1} + \beta_1 u_t + \beta_2 u_{t+1} + \dots + \beta_n u_{t+n-1} + d u_{t+n} \quad (\text{IV-50})$$

où les β_i et γ_i sont reliés par l'équation

$$\begin{bmatrix} 1 & & & & \\ -\alpha_n & & & & \\ & \ddots & & & \\ -\alpha_{n-1} & & & & \\ & \ddots & & & \\ -\alpha_2 & & \alpha_{n-1} & \alpha_n & 1 \end{bmatrix} \times \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = \begin{bmatrix} \beta_n \\ \vdots \\ \beta_1 \end{bmatrix} + d \begin{bmatrix} \alpha_n \\ \vdots \\ \alpha_1 \end{bmatrix} \quad (\text{IV-51})$$

Généralisons maintenant au système (IV-42), et pour cela introduisons au préalable les notations suivantes :

$$\underline{y}_t = \begin{bmatrix} y_t^1 \\ \vdots \\ y_t^p \end{bmatrix} ; \underline{x}_t = \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^p \end{bmatrix} ; \underline{x}_t^i = \begin{bmatrix} x_t^{i1} \\ \vdots \\ x_t^{in_i} \end{bmatrix} \underline{\Delta} \begin{bmatrix} x_t^{s_{i-1}+1} \\ \vdots \\ x_t^{s_i} \end{bmatrix} ; D = \begin{bmatrix} d_i^T \\ \vdots \\ d_i^T \end{bmatrix} ; \Gamma_i = \begin{bmatrix} \gamma_{i1}^T \\ \vdots \\ \gamma_{in_i}^T \end{bmatrix} \quad (\text{IV-52})$$

Compte tenu de la forme particulière des ϕ_{ij} (IV-43), la démarche, effectuée dans la cas monovariable, s'applique encore pour déterminer l'état $\underline{x}_t^i$ relatif au sous-système i .

Soit d'après (IV-48) :

$$\underline{x}_t^i = \begin{bmatrix} y_t^i \\ y_{t+1}^i \\ \vdots \\ y_{t+n_i-1}^i \end{bmatrix} - \begin{bmatrix} d_i^T & & & & \\ \gamma_{i1}^T & d_i^T & & & \\ & \ddots & \ddots & & \\ \gamma_{in_i-1}^T & & \gamma_{i1}^T & d_i^T & \\ & & & & \end{bmatrix} \times \begin{bmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+n_i-1} \end{bmatrix} \quad (\text{IV-53})$$

A la différence du cas précédent le terme ξ , qui représente ici $\underline{x}_{t+1}^i$, fait intervenir les couplages de Φ par l'intermédiaire du produit de la ligne s_i de Φ par le vecteur $[(\underline{x}_t^1)^T, \dots, (\underline{x}_t^i)^T]^T$, de sorte que l'on a :

$$y_{t+n_i}^i = \sum_{j=1}^i \alpha_{ij}^T \underline{x}_t^j + \gamma_{in_i}^T \underline{u}_t + \gamma_{in_i-1}^T \underline{u}_{t+1} + \dots + \gamma_{i1}^T \underline{u}_{t+n_i-1} + d_i^T \underline{u}_{t+n_i} \quad (IV-54)$$

où α_{ij}^T désigne la dernière ligne du bloc Φ_{ij} . Là, encore, cette expression peut être transformée en éliminant $\underline{x}_t^i$ à l'aide de (IV-53). Bien que simple dans son principe, la relation analogue à (IV-51) se complique singulièrement par la présence des couplages et en raison de la taille différente des blocs Φ_{ij} .

Considérons d'abord l'élimination de $\underline{x}_t^i$; en substituant (IV-53) dans (IV-54), il vient :

$$y_{t+n_i}^i = \sum_{j=1}^{i-1} \alpha_{ij}^T \underline{x}_t^j + \alpha_{i11} y_t^i + \dots + \alpha_{iin_i} y_{t+n_i-1}^i + \beta_{i1}^T \underline{u}_t + \dots + \beta_{in_i}^T \underline{u}_{t+n_i-1} + d_i^T \underline{u}_t \quad (IV-55)$$

avec la relation entre β_{ik} et γ_{ik} donnée par :

$$\begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ -\alpha_{i11} & & & & \\ & \ddots & & & \\ & & & & 1 \end{bmatrix} \times \begin{bmatrix} \gamma_{i1}^T \\ \vdots \\ \gamma_{in_i}^T \end{bmatrix} = \begin{bmatrix} \beta_{i1}^T \\ \vdots \\ \beta_{in_i}^T \end{bmatrix} + \begin{bmatrix} \alpha_{i11} d_i^T \\ \vdots \\ \alpha_{iin_i} d_i^T \end{bmatrix} \quad (IV-56)$$

premiers résultats que l'on peut considérer comme l'extension directe de (IV-50) et (IV-51).

L'élimination des états relatifs aux précédents sous-systèmes $(\underline{x}_t^j, j = 1, \dots, j = 1, \dots, i-1)$ se poursuit de la même manière en apportant par l'intermédiaire de (IV-53) leur contribution dans la pondération des entrées $\underline{u}_t$. La relation explicite entre les *nouveaux* β_{ik} et les paramètres de la réalisation γ_{ik} est exprimée à l'équation (IV-57) (voir tableau 1) en supposant pour clarifier la structure de l'équation matricielle que $n_p > n_1 > n_2$.

En toute généralité la relation entrée/sortie s'écrira :

$$y_{t+n_i}^i = \sum_{j=1}^{i-1} \sum_{k=1}^{n_j} \alpha_{ijk} y_{t+k-i}^j + \alpha_{i11} y_t^i + \dots + \alpha_{iin_i} y_{t+n_i-1}^i + \beta_{i1}^T \underline{u}_t + \dots + \beta_{in_i}^T \underline{u}_{t+n_i-1} + d_i^T \underline{u}_{t+n_i} \quad (IV-58)$$

où les termes associés aux Σ sont relatifs au couplage.

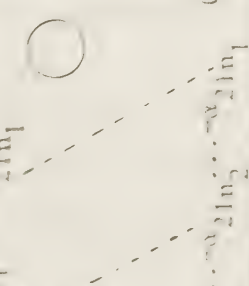
1 α_{11n_1} $\cdot$ $\cdot$ $\cdot$ α_{113} α_{112} α_{113} $\dots$ α_{11n_1} 1		T_{111} T_{112} $\cdot$ $\cdot$ T_{1n_1}	T_{1n_1} T_{1n_1+1} $\cdot$ $\cdot$ T_{111}	$\alpha_{11n_1} \frac{d}{d_1}$ $\alpha_{11n_1+1} \frac{d}{d_1}$ $\cdot$ $\cdot$ $\alpha_{111} \frac{d}{d_1}$
$\alpha_{21n_2+1} \dots \alpha_{21n_1}$ α_{21n_2} $\cdot$ $\cdot$ $\cdot$ $\alpha_{212} \dots \alpha_{21n_2} \dots \alpha_{21n_1}$		T_{221} $\cdot$ $\cdot$ $\cdot$ T_{2n_2}	T_{2n_2} $\cdot$ $\cdot$ $\cdot$ T_{221}	$\alpha_{22n_2} \frac{d}{d_2} + \alpha_{21n_2} \frac{d}{d_1}$ $\cdot$ $\cdot$ $\cdot$ $\alpha_{221} \frac{d}{d_2} + \alpha_{211} \frac{d}{d_1}$
$\cdot$ $\cdot$ $\cdot$		$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$
0 $\alpha_{p1n_1} \dots \alpha_{p1n_1}$ $\cdot$ $\cdot$ $\cdot$ $\alpha_{p12} \dots \alpha_{p1n_1}$		T_{p1} T_{p2} $\cdot$ $\cdot$ T_{pn_p}	T_{pn_p} T_{pn_p-1} $\cdot$ $\cdot$ $\cdot$ T_{p1}	$\alpha_{ppn_p} \frac{d}{d_p} + \sum_{j=1}^{p-1} \alpha_{pjn_p} \frac{d}{d_j}$ $\cdot$ $\cdot$ $\cdot$ $\alpha_{ppn_p-1} \frac{d}{d_p} + \sum_{j=1}^{p-1} \alpha_{pjn_p-1} \frac{d}{d_j}$ $\cdot$ $\cdot$ $\cdot$ $\alpha_{pp1} \frac{d}{d_p} + \sum_{j=1}^{p-1} \alpha_{pj1} \frac{d}{d_j}$ $\cdot$ $\cdot$ $\cdot$ α_{pn_p}

Tableau 1

Ces derniers résultats (IV-57) et (IV-58), d'une part, généralisent ceux de BONIVENTO et GUIDOZI [4] (introduction de D) et, d'autre part, les simplifient très significativement. En effet, sur le plan calcul, l'élimination, partielle ou totale, des états de la relation d'entrée/sortie (IV-54) est rendue triviale par la forme triangulaire de (IV-57), conséquence de la structure particulière des formes canoniques que nous avons développées. Cette propriété se retrouve explicitée dans (IV-58) où n'interviennent plus les sorties relatives du sous système suivant (i)

Passage de la forme (4)

Rappelons que

$$(1) \xrightarrow{P} (4) \quad \text{avec} \quad P = \begin{bmatrix} P_1 & O \\ O & P_p \end{bmatrix} \quad \text{et} \quad P_i = \begin{bmatrix} O & 1 \\ 1 & O \end{bmatrix}$$

Il est clair que la relation d'entrée/sortie (IV-58) reste inchangée. Pour le reste, il s'agit de permuter les états $\underline{x}_t^i$ et les lignes correspondantes dans les équations telles que (IV-53), (IV-56) et (IV-57), opérations effectuées au moyen d'une multiplication à gauche par P_i . A titre d'exemple nous ne réécrivons simplement que (IV-53) qui devient alors :

$$\underline{x}_t^i = \begin{bmatrix} y_{t+n_i-1}^i \\ \vdots \\ y_{t+1}^i \\ y_t^i \end{bmatrix} = \begin{bmatrix} Y_{in_i-1}^T & \cdots & Y_{i1}^T & d_i^T \\ \vdots & & & \\ Y_{i1} & & & O \\ d_i^T & & & \end{bmatrix} \times \begin{bmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+n_i-1} \end{bmatrix}$$

Remarque : Il convient de noter, dans l'un et l'autre cas, que les blocs diagonaux de la matrice qui intervient dans (IV-57) ne sont autres que les matrices de passage L_i . Ceci suggère certaines simplifications en faisant appel aux structures canoniques (1) et (3) que l'on aborde maintenant.

4-2. Relations d'entrée/sortie associées à la forme canonique observable (1)-(3)

Rappelons que ce second type de structure "observable" (prenons la forme (1)), se déduit de celle précédemment traitée (forme (2) par exemple) par un changement de base spécifique, noté L et défini par :

$$L = \begin{bmatrix} L_1 & \text{O} \\ \text{O} & L_p \end{bmatrix} \quad \text{avec } L_i = \begin{bmatrix} 1 & & & \\ -\alpha_{iin_i} & \ddots & & \\ & \ddots & \ddots & \\ -\alpha_{ii2} & & -\alpha_{iin_i} & 1 \end{bmatrix} \quad (\text{IV-59})$$

Posons alors :

$$\underline{n}_t = L \underline{x}_t \quad (\text{IV-60})$$

de sorte que le système décrit par (IV-42) aura pour nouvelle équation d'état :

$$\underline{n}_{t+1} = \Phi' \underline{n}_t + \Gamma' \underline{u}_t \quad (\text{IV-61})$$

$$\underline{y}_t = H' \underline{n}_t + D' \underline{u}_t$$

où d'une part :

$$\Phi' = L \Phi L^{-1} \quad \left\{ \begin{array}{l} \Phi'_{ii} = L_i \Phi_{ii} L_i^{-1} = \begin{bmatrix} \alpha_{iin_i} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \alpha_{ii1} & & & 0 \end{bmatrix} \\ \Phi'_{ij} = \Phi_{ij} L_j^{-1} = \begin{bmatrix} \text{O} \\ \vdots \\ \alpha'_{ij1} \dots \alpha'_{ijn_j} \end{bmatrix} \\ j=1,2,\dots,i-1 \end{array} \right. \quad \underbrace{\alpha'_{ij}^T}_{\alpha'_{ij}^T} \quad (\text{IV-62})$$

et d'autre part :

$$H' = H, \quad D' = D \quad (\text{IV-63})$$

et :

$$\Gamma' = \begin{bmatrix} B_1 \\ \vdots \\ B_2 \\ \vdots \\ B_p \end{bmatrix} ; \quad B_i = \begin{bmatrix} b_{i1}^T \\ \vdots \\ b_{in_i}^T \end{bmatrix} = L_i \begin{bmatrix} \gamma_{i1}^T \\ \vdots \\ \gamma_{in_i}^T \end{bmatrix} \quad (\text{IV-64})$$

Ces relations préliminaires étant données, reprenons directement l'expression (IV-53) en notant que de (IV-62), il vient :

$$\alpha_{ij}^{\prime T} = \alpha_{ij}^T L_j^{-1} \quad (\text{IV-65})$$

or, par définition de $\underline{n}_t$ et par analogie aux relations antérieures, on aura :

$$\underline{x}_t^j = L_j^{-1} \underline{n}_t^j$$

d'où :

$$\alpha_{ij}^T x_t^j = \alpha_{ij}^T L_j^{-1} n_t^j = \alpha_{ij}^T n_t^j$$

de sorte que (IV-55) s'écrit maintenant :

$$y_{t+n_i}^i = \sum_{j=1}^{i-1} \alpha_{ij}^T n_t^j + \alpha_{i11} y_t^i + \dots + \alpha_{iin_i} y_{t+n_i-1}^i + \beta_{i1}^T u_t + \dots + \beta_{in_i}^T u_{t+n_i-1} + d_i^T u_{t+n_i} \quad (IV-66)$$

La relation, entre β_{ik} et les paramètres de la réalisation, se déduit immédiatement de (IV-56), compte tenu de (IV-64), il vient :

$$\beta_{ik} = b_{in_i+1-k} - \alpha_{iik} d_i \quad (IV-67)$$

Explicitons maintenant n_t en fonction de y_t et u_t . A l'aide de (IV-64), on déduit immédiatement de (IV-53) l'expression :

$$n_t^i = L_i \times \begin{bmatrix} y_t^i \\ y_{t+1}^i \\ \vdots \\ y_{t+n_i-1}^i \end{bmatrix} - \begin{bmatrix} d_i^T & & & \\ b_{i1}^T - \alpha_{i11} d_i^T & & & \\ \vdots & \ddots & \ddots & \\ b_{in_i-1}^T - \alpha_{ii2} d_i^T & \dots & b_{i1}^T - \alpha_{iin_i} d_i^T & d_i^T \end{bmatrix} \begin{bmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+n_i-1} \end{bmatrix} \quad (IV-68)$$

que l'on peut encore exprimer (cf IV-67) :

$$n_t^i = - \begin{bmatrix} 1 & & & & \\ \alpha_{i11} & & & & \\ \vdots & \ddots & \ddots & \ddots & \\ \alpha_{ii3} & & & & \\ \alpha_{ii2} & \alpha_{ii3} & \dots & \alpha_{iin_i} & 1 \end{bmatrix} \begin{bmatrix} d_i^T & & & & \\ \beta_{i1}^T & d_i^T & & & \\ \vdots & \vdots & \ddots & \ddots & \\ \beta_{i3}^T & & & & \\ \beta_{i2}^T & \beta_{i3}^T & \dots & \beta_{in_i}^T & d_i^T \end{bmatrix} \begin{bmatrix} y_t^i \\ \vdots \\ y_{t+n_i-1}^i \\ \hline u_t \\ \vdots \\ u_{t+n_i-1} \end{bmatrix} \quad (IV-69)$$

Il est clair que la substitution de (IV-69) dans (IV-66) redonne l'expression (IV-58) ; par suite la contre partie de (IV-57) ne présente aucun intérêt.

Passage à la forme canonique (3)

Là encore on a la transformation :

$$(1) \xrightarrow{P} (3) \quad \text{avec} \quad P = \begin{bmatrix} P_1 & O \\ O & P_p \end{bmatrix} \quad \text{et} \quad P_i = \begin{bmatrix} O & & 1 \\ \vdots & \ddots & \vdots \\ 1 & & O \end{bmatrix}$$

qui consiste à permuter individuellement les composantes des états n_t^i . Bien évidemment la relation d'entrée/sortie (IV-58) demeure inchangée. Quant à l'équation (IV-66) elle est toujours vérifiée en notant que les composantes de α_{ij}^T (cf ch III§6) ont été permutées de la même manière que celle de n_t^j de

sorte que leur produit demeure identique. L'expression (IV-67) ne subit aucune modification. Enfin (IV-69) devient, après prémultiplication par P_i en notant $\underline{n}_t^{i'}$ l'état permuté :

$$\underline{n}_t^{i'} = - \left[\begin{array}{c|c} \alpha_{ii2} \alpha_{ii3} \dots \alpha_{iin_i} & \begin{matrix} \beta_{i2}^T & \beta_{i3}^T & \dots & \beta_{in_i}^T & d_i^T \end{matrix} \\ \alpha_{ii3} & \beta_{i3}^T \\ \vdots & \vdots \\ \alpha_{iin_i} & \beta_{in_i}^T \\ 0 & d_i^T \end{array} \right] \times \begin{bmatrix} y_t^i \\ \vdots \\ y_{t+n_i-1}^i \\ \underline{u}_t \\ \vdots \\ \underline{u}_{t+n_i-1} \end{bmatrix} \quad (IV-70)$$

... résultats que nous avons établis différemment dans [5] (équations (1) et (2))

En résumé, étant donné un modèle d'état ayant la structure canonique observable, on peut lui associer une relation d'entrée/sortie ne dépendant que des invariants du système (IV-58) et de paramètres dont se déduisent aisément les matrices Γ et D de la forme (2) (Bucy) ou (4) (IV-57). On a également vu que le système triangulaire (IV-57) pouvait être découplé en faisant intervenir dans la relation d'entrée/sortie les états relatifs aux sous systèmes mono-sortie déjà traité, (IV-53-54 et 56). Rappelons enfin qu'il en découle une simplification entre la paramétrisation de la relation d'entrée/sortie et celle du modèle d'état, particulièrement significative en ce qui concerne les deux autres formes canoniques (1) et (3) dont la détermination est alors immédiate (IV-66, 67 et 69).

Avant d'aborder la phase proprement dite d'identification des paramètres n_i et α_{ijk} , il convient de préciser les deux points suivants.

4-3. Identifiabilité et Réalisabilité

La présence explicite d'une entrée $\underline{u}_t$ est l'élément nouveau qui distingue ce problème de réalisation des deux précédents. Il est clair, tout au moins intuitivement, que celle-ci doit "exciter" suffisamment le système pour en permettre une identification dynamique. On connaît également les difficultés que peut représenter un retour de la sortie vers l'entrée. Ces considérations nous amènent donc à parler d'*identifiabilité* dans les termes suivants.

Définition IV-1

Etant donné un système irréductible, d'ordre n , ce dernier sera *identifiable* à partir d'un échantillon d'entrées/sorties si, et seulement si, on peut construire à partir de ces entrées/sorties une réalisation de ce système.

La notion d'identifiabilité, telle que nous venons de l'introduire ici, est en fait un préalable au concept développé par exemple par GLOVER et WILLEMS [6] dans un contexte déterministe, et par E. TSE et al [7][8] d'un point de vue stochastique. Ces auteurs posent, en effet, le problème d'unicité du modèle, étant entendu que le système peut être effectivement identifié à partir de données.

Les développements qui suivent sont plutôt à rapprocher de ceux relatifs à la formulation probabiliste d'ASTROM [9], plus couramment en usage. Ils porteront donc sur la caractérisation d'entrées permettant d'identifier le système, l'unicité de la solution découlant de la paramétrisation du modèle.

Cette terminologie étant précisée, considérons maintenant la question corollaire de la *réalisabilité*, déjà introduite dans le cadre de la réalisation d'un modèle externe, et qui, rappelons-le, exprime la possibilité de modéliser un système par une réalisation de dimension *finie*. On verra, là encore, qu'il est théoriquement nécessaire de disposer d'une suite infinie de couples entrées/sorties, pour pouvoir répondre à la question. Cependant, à la différence de la réalisabilité d'une suite de paramètres de MARKOV, l'hypothèse d'une entrée insuffisamment excitante pourra conduire à une réalisation minimale de dimension inférieure à la réalité. En d'autres termes une réalisation minimale partielle, relative à un nombre fini de couples entrées/sorties ne sera caractéristique du système que si l'entrée $\{u_t\}$ vérifie certaines conditions, dites ici *d'identifiabilité*.

Ce sont celles-ci que nous développons maintenant en détail dans le cas monovariable, avant de les extrapoler successivement aux systèmes mono-sortie puis multivariables.

4-3-1. Système monovariable ($m = p = 1$)

Soit donc le système dont la transmittance irréductible d'ordre n :

$$F(z) = \frac{N(z)}{D(z)} = \frac{b_0 z^n + b_1 z^{n-1} + \dots + b_n}{z^n + a_1 z^{n-1} + \dots + a_n}$$

est entièrement définie par les $2n + 1$ paramètres a_i et b_j . On sait également (cf § 3) que ce système est complètement caractérisé par les $2n + 1$ premiers points de sa séquence de pondération $\{f_0, f_1, \dots, f_{2n}\}$

Par suite, en se plaçant dans l'hypothèse où le système est initialement au repos, son identification pourra être effectuée en résolvant l'une ou l'autre des équations suivantes :

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{2n} \end{bmatrix} = \begin{bmatrix} u_0 & & & \\ u_1 & u_0 & & \\ \vdots & \vdots & \ddots & \\ u_{2n} & & & u_0 \end{bmatrix} \times \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{2n} \end{bmatrix} \quad (\text{IV-71})$$

que l'on note :

$$\underline{y} = U \underline{f}$$

ou

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \\ \vdots \\ y_{2n} \end{bmatrix} = \begin{bmatrix} \xrightarrow{n+1} & \xrightarrow{n} \\ u_0 & 0 \\ u_1 & u_0 & y_0 \\ \vdots & \vdots & \vdots \\ u_n & \vdots & y_{n-1} & y_0 \\ \vdots & \vdots & \vdots & \vdots \\ u_{2n} & \vdots & y_{2n-1} & y_n \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \\ \vdots \\ a_1 \\ \vdots \\ a_n \end{bmatrix} \quad (\text{IV-72})$$

soit :

$$\underline{y} = R \underline{\alpha}$$

Désignons maintenant par $\mathcal{H}$ la matrice

$$\mathcal{H} = \begin{bmatrix} 1 & & 0 & \\ & \ddots & f_0 & \\ & & \vdots & \\ & & f_{n-1} & f_0 \\ & & f_n & f_1 \\ & & \vdots & \vdots \\ & & f_{2n-1} & f_n \end{bmatrix} \quad \begin{matrix} \xrightarrow{n+1} & \xrightarrow{n} \\ \uparrow & \uparrow \\ 2n+1 \end{matrix}$$

On vérifie alors facilement la relation

$$\mathcal{H} \underline{\alpha} = \underline{f} \quad (\text{IV-73})$$

et en prémultipliant à gauche par U , il vient :

$$U \mathcal{H} \underline{\alpha} = U \underline{f}$$

que l'on identifie à (IV-72), c'est-à-dire :

$$U \mathcal{H} = R \quad (\text{IV-74})$$

A ces système (IV-71) et (IV-72) ne correspond une solution unique en $\underline{h}$ ou $\underline{\alpha}$ que si, et seulement si, l'on a respectivement : $\rho(U)=2n+1$ ou $\rho(R) = 2n+1$. Or il est clair que la matrice $\mathcal{H}$ est régulière, ces deux blocs diagonaux l'étant, par suite, il vient d'après (IV-74) :

$$\rho(U) = 2n+1 \iff \rho(R) = 2n+1$$

Enfin compte tenu de la forme triangulaire de U, on énoncera la ...

Proposition IV-1

Etant donné un système monovariable d'ordre n , initialement au repos, celui-ci sera identifiable à partir de la suite (u_t, y_t) $t = 0, 1, \dots, N$, si, et seulement si, $u_0 \neq 0$ et $N > 2n$.

Dans le cas plus fréquent en pratique, où le système se trouve dans des conditions initiales quelconques, le problème est plus délicat, comme on va le voir.

Exemple et contre-exemple

Considérons le système du premier ordre :

$$y_t = b_0 u_t + b_1 u_{t-1} + a y_{t-1}$$

et écrivons cette relation, sous forme matricielle, pour $t = 1, 2, 3$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} u_1 & u_0 & y_0 \\ u_2 & u_1 & y_1 \\ u_3 & u_2 & y_2 \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ a \end{bmatrix} \quad (\text{IV-75})$$

Simulons un cas particulier en prenant $a = b_0 = b_1 = 1$, et appliquons la séquence d'entrée $u_0 = 1, u_1 = 0, u_2 = 1, u_3 = 0$. On en déduit immédiatement :

$$\begin{cases} y_1 = 1 + y_0 \\ y_2 = 1 + y_1 = 2 + y_0 \end{cases}$$

Substituons dans (IV-75), il en découle la relation :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & y_0 \\ 1 & 0 & 1+y_0 \\ 0 & 1 & 2+y_0 \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ a \end{bmatrix} \quad (\text{IV-76})$$

$$\underline{A} \quad R \quad \underline{\alpha}$$

On obtient immédiatement :

$$-\det R = 2+y_0-y_0 = 2 \neq 0 \quad \forall y_0$$

En conséquence, quelle que soit la condition initiale y_0 , c'est-à-dire quelle que soit l'entrée passée, la séquence choisie u_0, u_1, u_2, u_3 permettra d'identifier le système.

Dans ce cas la condition nécessaire (évidente d'après IV-75)

$$\rho \begin{bmatrix} u_0 & u_1 \\ u_1 & u_2 \\ u_2 & u_3 \end{bmatrix} = n+1 = 2$$

... semble donc également suffisante.

Simulons maintenant un second système défini par $a = \frac{1}{2}$ et $b_0 = b_1 = 1$. En utilisant la même entrée, il vient :

$$\begin{cases} y_1 = 1 + \frac{1}{2} y_0 \\ y_2 = 1 + \frac{1}{2} y_1 = \frac{3}{2} + \frac{1}{4} y_0 \\ y_3 = 1 + \frac{1}{2} y_2 = \frac{7}{4} + \frac{1}{8} y_0 \end{cases}$$

Substituons encore une fois dans (IV-75), on en déduit :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & y_0 \\ 1 & 0 & 1 + \frac{1}{2} y_0 \\ 0 & 1 & \frac{3}{2} + \frac{1}{4} y_0 \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ a \end{bmatrix} \quad (\text{IV-77})$$

$$\Delta \quad R \quad a$$

Cette fois on a :

$$-\det R = \frac{3}{2} - \frac{3}{4} y_0$$

$$\text{soit } \det R = 0 \quad \text{si } y_0 = 2$$

Donc pour une certaine entrée passée conduisant à $y_0 = 2$, la séquence d'entrée choisie ne permettra pas d'identifier le système. On constatera par ailleurs que l'on a $y_0 = y_1 = y_2 = y_3 = 2$! Introduisons maintenant une quatrième équation en posant $u_4 = v$, de sorte que (IV-77) devient :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & y_0 \\ 1 & 0 & 1 + \frac{1}{2} y_0 \\ 0 & 1 & \frac{3}{2} + \frac{1}{4} y_0 \\ 1 & 0 & \frac{7}{4} + \frac{1}{8} y_0 \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ a \end{bmatrix} \quad (\text{IV-78})$$

Si l'on cherche à déterminer les paramètres du modèle à l'aide de trois quelconques des quatre équations (IV-78), on vérifie facilement que l'on doit prendre $v \neq 0$, lorsqu'on se place dans la situation particulière $y_0 = 2$. Cette condition est précisément celle pour laquelle :

$$\rho \begin{bmatrix} u_0 & u_1 & u_2 \\ u_1 & u_2 & u_3 \\ u_2 & u_3 & u_4 \end{bmatrix} = 3 = 2n+1 \quad (\text{IV-79})$$

Voyons pourquoi. Pour cela notons $y_t(y_0)$ l'application linéaire exprimant la sortie à l'instant t en fonction du paramètre y_0 . Reprenons les trois premières équations de (IV-76), puis les trois dernières, auxquelles on associe respectivement les matrices R_1 et R_2 :

$$R_1 = \begin{bmatrix} u_1 & u_0 & y_0 \\ u_2 & u_1 & y_1(y_0) \\ u_3 & u_2 & y_2(y_0) \end{bmatrix} \quad R_2 = \begin{bmatrix} u_2 & u_1 & y_1(y_0) \\ u_3 & u_2 & y_2(y_0) \\ u_4 & u_3 & y_3(y_0) \end{bmatrix}$$

... que l'on regroupe pour former l'unique tableau

$$\begin{bmatrix} y_1(y_0) & u_2 & u_1 & u_0 & y_0 \\ y_2(y_0) & u_3 & u_2 & u_1 & y_1(y_0) \\ y_3(y_0) & u_4 & u_3 & u_2 & y_2(y_0) \end{bmatrix}$$

Par définition, (Cf équation (IV-75)), la colonne 1 est combinaison linéaire des colonnes 3, 4 et 5, résultat que nous notons :

$$(1) = \alpha(3) + \beta(4) + \gamma(5) \quad (\text{IV-80})$$

Admettons que y_0 soit tel que le déterminant de R_1 soit nul. Ceci implique :

$$(5) = \delta(3) + \epsilon(4)$$

En substituant dans (IV-80), il vient :

$$(1) = \alpha'(3) + \beta'(4)$$

Si la condition (IV-70) est vérifiée, on en déduit que :

$$\begin{aligned} \nexists \alpha'' \text{ et } \beta'' : (2) &= \alpha''(3) + \beta''(4) \\ \Rightarrow \nexists \alpha''' \text{ et } \beta''' : (1) &= \alpha'''(2) + \beta'''(3) \end{aligned}$$

En d'autres termes, la condition (IV-79) est suffisante pour pouvoir identifier notre système.

Considérons maintenant le cas général d'un système d'ordre n , tel que :

$$y_t = a_1 y_{t-1} + \dots + a_n y_{t-n} + b_0 u_t + b_1 u_{t-1} + \dots + b_n u_{t-n} \quad (\text{IV-81})$$

Définissons alors la matrice R_k^ℓ par :

$$R_k^\ell = \left[\begin{array}{cccc|cccc} u_0 & u_1 & \cdot & \cdot & u_\ell & y_0 & y_1 & \cdot & \cdot & y_\ell \\ u_1 & u_2 & \cdot & \cdot & u_{\ell+1} & y_1 & y_2 & \cdot & \cdot & y_{\ell+1} \\ \cdot & \cdot & & & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot & & & \cdot \\ u_K & u_{K+1} & \cdot & \cdot & u_{K+\ell} & y_K & y_{K+1} & \cdot & \cdot & y_{K+\ell} \end{array} \right] \quad k+1 \quad (\text{IV-82})$$

$\xrightarrow{\ell+1} \qquad \qquad \qquad \xrightarrow{\ell+1}$

que l'on note encore :

$$R_k^\ell \triangleq [U_k^\ell \mid Y_k^\ell] \quad (\text{IV-83})$$

et également :

$$R_k^\ell \triangleq [\tilde{R}_k^\ell \mid y_\ell] \quad (\text{IV-84})$$

où y_ℓ est la dernière colonne de R_k^ℓ

D'après (IV-81), pour tout $k : k \geq 2n$, n doit être la plus petite valeur de ℓ telle que y_ℓ soit une combinaison linéaire exacte des colonnes $\tilde{R}_k^\ell$, sinon le système serait d'ordre inférieure à n . Mais cette dépendance linéaire n'est caractéristique de l'ordre du système que si $\tilde{R}_k^n$ est de rang maximum $2n+1$. Il s'agit donc de déterminer les conditions sur $\{u_t\}$ pour qu'il en soit ainsi.

Définition IV-2

On dira que la séquence d'entrée u_t , $t = 0, 1, \dots, N > 4q$ offre une identifiabilité d'ordre q si, et seulement si :

$$\left[\begin{array}{ccc} u_0 & u_1 & u_{2q} \\ u_1 & u_2 & \cdot \\ \cdot & \cdot & \cdot \\ u_{N-2q} & \cdot & u_N \end{array} \right] = 2q + 1 \quad (\text{IV-85})$$

résultat que l'on note encore : $\rho(U_{N-2q}^{2q}) = 2q+1$

Théorème IV-2

Etant donné un système monovarié d'ordre n , celui-ci sera identifiable à partir de la séquence $\{u_t, y_t\}$ $t = 0, 1, \dots, N$ avec $N \geq 4n$, si l'entrée $\{u_t\}$ offre une identifiabilité d'ordre n .

Démonstration ($N = 4n$)

Introduisons la matrice R_r^* , $r = 0, 1, 2 \dots$ définie comme suit :

$$R_r^* = \left[\begin{array}{ccc|ccc} u_r & \cdot & \cdot & u_{r+n} & y_r & \cdot & \cdot & y_{r+n-1} \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \hline u_{r+2n} & \cdot & \cdot & u_{r+3n} & y_{r+2n} & \cdot & \cdot & y_{r+3n-1} \end{array} \right] \begin{array}{c} \updownarrow \\ 2n+1 \end{array} \quad (IV-86)$$

$\xleftarrow{\quad n+1 \quad} \quad \quad \quad \xleftarrow{\quad n \quad}$

qui représente en fait $2n+1$ lignes consécutives de la matrice $\tilde{R}_{N-n}^*$. Nous noterons également $\mathcal{U}_r$ l'espace engendré par les $n+1$ premières colonnes de R_r^* . Supposons maintenant (IV-85) satisfait, et plaçons nous dans la situation la plus défavorable correspondant à des conditions initiales $y_0, y_1, \dots, y_{n-1}$ telles que : $\rho(R_0^*) = n+1$. D'après (IV-81) la dernière colonne R_1^* , notée y_n^* , est combinaison linéaire des $2n+1$ colonnes de R_0^* ; mais compte tenu de l'hypothèse précédente, y_n^* peut être exprimé uniquement en fonction de $n+1$ premières colonnes de R_0^* . En conséquence y_n^* appartient à $\mathcal{U}_0$, mais par ailleurs n'appartient pas à $\mathcal{U}_r$, $r \geq 1$ en raison de (IV-85). Poursuivons ce raisonnement avec y_{n+1}^* qui par définition, équation (IV-81), est combinaison linéaire des colonnes de R_1^* . En vertu de ce qui précède, on en déduit que : $y_{n+1}^* \in \mathcal{U}_0 \cup \mathcal{U}_1$, mais là encore : $y_{n+1}^* \notin \mathcal{U}_r$, $r \geq 2$. Plus généralement, on vérifie que : $y_{n+r}^* \in \bigcup_{i=0}^r \mathcal{U}_i$ et $y_{n+r}^* \notin \mathcal{U}_i$, $i > r$. Donc finalement :

$y_{2n-1}^* \in \bigcup_{i=0}^{n-1} \mathcal{U}_i$ et $y_{2n-1}^* \notin \mathcal{U}_n$. En d'autres termes $y_n^*, y_{n+1}^*, \dots, y_{2n-1}^*$ sont linéairement indépendants et n'appartiennent pas à $\mathcal{U}_n$, de sorte qu'on aura nécessairement $\rho(R_n^*) = 2n+1$. Le système sera donc identifiable.

Remarque : Si $N > 4n$, il suffit de raisonner avec des matrices R_r^* ayant $N-2n+1$ lignes.

Il est clair que le fait de disposer d'un échantillon de longueur N supérieur à $4n$ n'apporte ici aucune information supplémentaire sur le plan

théorique. En revanche dans le cas pratique (problème d'approximation, cf Ch. 5) où les systèmes d'équations surabondants seront résolus au sens des moindres carrés, il est particulièrement indiqué (pour des raisons évidentes) que l'entrée permette continuellement d'identifier le système. C'est cette notion que nous précisons dans la...

Définition IV-3

La séquence d'entrée $\{u_t\}$, $t = 0, 1, \dots, N > 4q$ présente une excitation permanente d'ordre q si, et seulement si, pour $p = 2q$, on a :

$$\rho \begin{bmatrix} u_r & \cdot & \cdot & u_{r+2q} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ u_{r+p} & \cdot & \cdot & u_{r+p+2q} \end{bmatrix} = 2q+1, \quad r = 0, \dots, N-p-2q$$

... qu'il est intéressant de rapprocher du concept d'excitation persistante (Astrom [9]). Celui-ci se place, rappelons-le dans l'hypothèse N infini de sorte que l'idée de permanence impliquée par $p = 2q$, pour N fini, est conservée si la condition précédente de rang n est vérifiée, quel que soit r , que pour un certain $p > 2q$, mais fini. Par suite si l'on admet que les deux limites suivantes existent :

$$\bar{u} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u_t \quad \text{et} \quad s_\tau = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u_t u_{t+\tau}$$

la notion d'excitation permanente (avec p fini) et la notion d'excitation persistante s'impliquent mutuellement. En effet introduisons la matrice

$$Q = \begin{bmatrix} s_0 & \cdot & \cdot & s_{2q} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ s_{2q} & \cdot & \cdot & s_0 \end{bmatrix}$$

il est alors clair que si $\{u_t\}$ offre une excitation permanente d'ordre q , on a :

$$Q = \lim_{N \rightarrow \infty} \frac{1}{N} (U_{N-2q}^{2q})^T (U_{N-2q}^{2q}) > 0$$

Revenons maintenant au théorème IV-2, dont découle les deux conséquences suivantes :

Corollaire IV-1

Etant donné un système monovariable d'ordre fini, inférieur ou égal à q , identifiable à partir de la séquence d'entrées/sorties $\{u_t, y_t\}$ $t = 0, 1, \dots, N$

avec $N \geq 4q$, l'ordre n de ce système est alors donné par la première valeur de ℓ pour laquelle R_{N-q}^ℓ ne sera plus de rang maximal.

Remarque : Ceci suppose implicitement que l'entrée $\{u_t\}$ offre une identifiabilité d'ordre q . Par ailleurs quelle que soit la valeur de n ($< q$) on aura toujours $\rho(R_{N-q}^q) = q+n+1$

Corollaire IV-2

Un système monovarié sera réalisable si, et seulement si, il existe une valeur finie de ℓ , pour laquelle R_∞^ℓ n'est plus de rang maximal, l'entrée $\{u_t\}$ offrant a priori une identifiabilité d'ordre infini.

4-3-2. Système monosortie ($p=1$)

Nous nous intéressons ici directement à l'équivalent du théorème IV-2. On notera $\underline{u}_t$ le vecteur d'entrée de dimension m , appliqué au système défini par sa transmittance :

$$F(z) = \frac{\underline{b}_0^T z^n + \underline{b}_1^T z^{n-1} + \dots + \underline{b}_n^T}{z^n + \dots + a_n}$$

où les $\underline{b}_i$ sont des vecteurs de dimension m et n est le degré du P.P.C.M. des dénominateurs des m transmittances, relatives à chacune des entrées, et supposées irréductibles. Par suite on a la relation suivante, analogue à (IV-81) :

$$y_t = a_1 y_{t-1} + \dots + a_n y_{t-n} + \underline{b}_0^T \underline{u}_t + \underline{b}_1^T \underline{u}_{t-1} + \dots + \underline{b}_n^T \underline{u}_{t-n} \quad (IV-87)$$

On définirait de même des matrices R_k^ℓ , U_k^ℓ , Y_k^ℓ , $\tilde{R}_k^\ell$ et $\tilde{Y}_k^\ell$ en remplaçant u_t par $\underline{u}_t^T$ dans (IV-82), (IV-83), et (IV-84). En ce qui concerne la définition (IV-2) la condition (IV-85) devient :

$$\rho \left[\begin{array}{cccc} \underline{u}_0^T & \underline{u}_1^T & \dots & \underline{u}_{2q}^T \\ \vdots & \vdots & \ddots & \vdots \\ \underline{u}_{N-2q}^T & \dots & \dots & \underline{u}_N^T \end{array} \right] = (2q+1)m \iff \rho(U_{N-2q}^{2q}) = (2q+1)m \quad (IV-88)$$

$\xleftarrow{\hspace{1.5cm}}$
 $(2q+1)m$

N devant maintenant implicitement vérifier l'inégalité :

$$N \geq 2q(m+1) + m-1 \triangleq N(q) \quad (IV-89)$$

de façon à ce que la matrice U dans (IV-88) ait au moins autant de lignes que de

colonnes. On généraliserait de la même manière la définition IV-3 que l'on pourrait rapprocher dans les mêmes conditions de la formulation multivariable de l'excitation persistante [25].

Ceci étant précisé, le théorème IV-2 est encore exact, l'inégalité sur N étant maintenant $N \geq N(n)$.

La démonstration suit point par point celle du cas monovariable, en définissant cette fois la matrice R_r^* par :

$$R_r^* = \left[\begin{array}{ccc|ccc} \underline{u}_r^T & \cdot & \cdot & \underline{u}_{r+n}^T & y_r & \cdot & \cdot & y_{r+n-1} \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \underline{u}_{r+l}^T & \cdot & \cdot & \underline{u}_{r+l+n}^T & y_{r+l} & \cdot & \cdot & y_{r+l+n-1} \end{array} \right] \begin{array}{c} \uparrow \\ (2n+1)m \\ \downarrow \end{array} \quad (IV-90)$$

$\xleftarrow{(n+1)m} \quad \xrightarrow{n}$

avec $l = (2n+1)m-1$

En conséquence les corollaires (IV-1) et (IV-2) s'appliquent encore aux systèmes multi-entrées, en remplaçant toutefois la borne inférieure de N par l'expression (IV-89), là où elle intervient. Enfin, on a bien sûr $\rho(R_{N-q}^q) = (q+1)m+n$.

Remarque : En posant $m = 1$, on retrouve donc exactement les résultats directement développés dans l'hypothèse monovariable.

4-3-3. Système multivariable

Considérons un système dynamique multivariable Σ_n , irréductible et de structure $(n_1, n_2, \dots, n_p)$ pour l'instant supposée connue. Nous admettrons que l'entrée $\{\underline{u}_t\}$ offre une identifiabilité d'ordre n , ce qui suppose $N \geq N(n)$, cf équation (IV-89). Introduisons alors les notations suivantes :

$$R_\ell^i = \left[\begin{array}{ccc|ccc} \underline{u}_0^T & \cdot & \cdot & \underline{u}_\ell^T & y_0^i & \cdot & \cdot & y_\ell^i \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & & & \cdot \\ \underline{u}_{N'}^T & \cdot & \cdot & \underline{u}_{N'+\ell}^T & y_{N'}^i & \cdot & \cdot & y_{N'+\ell}^i \end{array} \right] \quad (IV-91)$$

avec : $N' = N - n$

puis :

$$Y^i = \left[\begin{array}{ccc} y_0^i & \cdot & \cdot & y_{n_i-1}^i \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ y_{N'}^i & \cdot & \cdot & y_{N'+n_i-1}^i \end{array} \right] \quad i = 1, 2, \dots, p \quad (IV-92)$$

et enfin :

$$\begin{cases} S_{\ell}^i = [Y^1 \mid Y^2 \mid \dots \mid Y^{i-1} \mid R_{\ell}^i] \\ S_{\ell}^1 = R_{\ell}^1 \end{cases} \quad (IV-93)$$

On définira alors comme précédemment les partitions :

$$R_{\ell}^i \triangleq [\tilde{R}_{\ell}^i \mid \underline{y}_{\ell}^i] \quad (IV-94)$$

et

$$S_{\ell}^i \triangleq \begin{array}{c} [\tilde{S}_{\ell}^i \mid \underline{y}_{\ell}^i] \\ \longleftrightarrow \quad \longleftrightarrow \\ v_i \quad 1 \end{array} \quad (IV-95)$$

avec

$$\begin{cases} v_i = s_{i-1} + (n_i + 1)m + n_i \\ s_i = \sum_{j=1}^i n_j, \quad s_0 = 0 \end{cases} \quad (IV-96)$$

Si l'on décompose ce système multivariable en p sous systèmes monosortie, on peut associer à chacun d'eux un modèle irréductible de degré $n'_1, n'_2, \dots, n'_p$ avec $n'_1 = n_1$ et $n \geq n'_i \geq n_i$ $i = 2, \dots, p$. Par suite, si l'entrée $\{\underline{u}_t\}$ offre une identifiabilité d'ordre n , chaque sous-système monosortie sera individuellement identifiable ; en conséquence le système original multivariable le sera.

Corollaire IV-3

Etant donné un système multivariable irréductible Σ_n de structure $(n_1, \dots, n_p)$, identifiable à partir de la séquence $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, \dots, N$, alors on a :

$$\rho(\tilde{S}_{n_i}^i) = v_i \quad (IV-97)$$

En effet le système étant identifiable, les matrices $R_{n_i}^i$, $i = 1, \dots, p$ sont de rang maximal. On peut donc construire d'une manière unique des réalisations irréductibles $\Sigma_{n_1}^*, \dots, \Sigma_{n_p}^*$, relatives à chacune des sorties. Par somme directe de ces réalisations, on obtient une réalisation du système multivariable généralement réductible mais définie d'une façon unique à partir des données $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, \dots, N$. En se référant aux résultats du chapitre III, on ne peut déduire de cette réalisation réductible qu'un seul ensemble d'invariants n_i, α_{ijk} . Or les paramètres α_{ijk} sont précisément d'après (IV-58) les coefficients exprimant la dépendance linéaire de $\underline{y}_{n_i}^i$ par rapport aux colonnes de $\tilde{S}_{n_i}^i$. En con-

séquence cette matrice ne peut être que de rang maximum, c'est-à-dire v_i sinon l'hypothèse d'identifiabilité ne serait pas vérifiée.

De ce raisonnement découle également le ...

Corollaire IV-4

Etant donné un système multivariable irréductible d'ordre au plus égale à q , identifiable à partir de la séquence $\{\underline{u}_t, \underline{y}_t\} t=0,1,\dots, N$; alors les paramètres $n_1, n_2, \dots, n_p$, définissant sa structure sont déterminés comme étant respectivement la première valeur de ℓ tel que $S_\ell^i (i=1,2,\dots, p)$ ne soit plus de rang maximum.

Remarque : La séquence $\{\underline{u}_t\}$ offre donc a priori une identifiabilité d'ordre q et l'on a bien sûr $N \geq N(q)$ (cf (IV-89).

... et le ...

Corollaire IV-5

Un système multivariable sera réalisable si, et seulement si, il existe des valeurs finies de ℓ pour lesquelles S_ℓ^i n'est plus de rang maximum ($i = 1, \dots, p$), l'entrée $\{\underline{u}_t\}$ offrant a priori une identifiabilité d'ordre infini.

4-4. Algorithmes de réalisation

Nous allons présenter successivement deux algorithmes conduisant respectivement à la forme canonique observable (2) (forme de Bucy) et à la forme canonique observable (1).

Rappelons que, d'une part les deux autres structures observables (3) et (4) se déduisent des précédentes par des permutations des états, et d'autre part l'algorithme développé en [5] et [10] conduit directement à la forme (3) dans l'hypothèse où $D \equiv 0$.

Nous admettrons ici que :

- le système est réalisable, donc que la dimension d'une réalisation irréductible est bornée supérieurement par une quantité q , ici supposée connue.
- le système à identifier est complètement accessible et observable
- le système est identifiable à partir de la séquence $\{\underline{u}_t, \underline{y}_t\} t = 0, 1, \dots, N$.

Dans ces conditions, des résultats du paragraphe 4-3 découlent d'une façon naturelle les deux procédures suivantes :

Algorithme 5

- 1 - Etant donné la suite $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, 1, \dots, N$; poser $N' = N-q$.
- 2 - Déterminer successivement pour $i = 1, 2, \dots, p$ la plus petite valeur de ℓ telle que $\underline{y}_\ell^i$ soit combinaison linéaire des colonnes de S_ℓ^i (Définitions en (IV-91) à (IV-95).
En se référant aux équations (IV-58) correspondantes, en déduire les paramètres $n_i, \alpha_{ijk}, \beta_{ik}$ et $\underline{d}_i$
- 3 - Résoudre le système triangulaire (IV-57) par rapport aux $\underline{\gamma}_{ik}$
- 4 - Construire la réalisation irréductible (Φ, Γ, H, D) suivant la forme définie en (IV-42) et (IV-43) que nous rappelons ici pour mémoire :

$$\left\{ \begin{array}{l} \Phi = \begin{bmatrix} \Phi_{11} & \text{---} & \text{---} \\ \Phi_{21} & \text{---} & \text{---} \\ \vdots & \vdots & \vdots \\ \Phi_{p1} & \text{---} & \text{---} \end{bmatrix} \quad \text{avec} \quad \Phi_{ij} = \begin{bmatrix} \text{---} & \text{---} & \text{---} \\ \vdots & \vdots & \vdots \\ \alpha_{ij1} & \dots & \alpha_{ijn_j} \end{bmatrix} \quad i > j \quad \text{et} \quad \Phi_{ii} = \begin{bmatrix} 0 & 1 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{ii1} & \dots & \alpha_{iin_i} & 1 \end{bmatrix} \\ \\ \Gamma = \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_p \end{bmatrix} \quad \text{avec} \quad \Gamma_i = \begin{bmatrix} \underline{\gamma}_{i1}^T \\ \vdots \\ \underline{\gamma}_{in_i}^T \end{bmatrix} ; \quad H = \begin{bmatrix} \text{---} & \text{---} & \text{---} & 1 & \text{---} & \text{---} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & 0 \end{bmatrix} \quad \begin{array}{l} \text{colonne } s_{i-1}+1 \\ \text{ligne } i \end{array} ; \quad D = \begin{bmatrix} \underline{d}_1^T \\ \vdots \\ \underline{d}_p^T \end{bmatrix}$$

- 5 - Calculer l'état initial $\underline{x}_0$ en utilisant successivement l'équation (IV-57) pour $i = 1, \dots, p$, en prenant $t = 0$.

Remarque : Comme toujours si un n_i est nul la ligne de couplage α_{ijk} relative aux matrices $\Phi_{i1}, \dots, \Phi_{i,i-1}$ prend place dans les s_i premières positions de la ligne i de la matrice H .

Démonstration - Elle découle directement des hypothèses concernant le système et son identifiabilité, et du corollaire IV-4.

A propos de ces hypothèses, notons cependant lorsque le processus à identifier n'est pas complètement accessible et observable, la réalisation obtenue, nécessairement observable, sera de dimension minimale compatible avec les données mais généralement pas accessible. Cependant si la partie non accessible mais observable de l'état initial est nulle la réalisation obtenue sera néanmoins irréductible. Il en sera de même, tout au moins asymptotiquement ($N \rightarrow \infty$) lorsque à défaut d'être nuls, ces états correspondent à des pôles stables.

Remarques :

. A la différence de l'algorithme 5 la construction, au cours de l'étape -2- des matrices S_{ℓ}^i nécessite le calcul de l'état relatif au sous système qui vient d'être traité, par l'intermédiaire de l'équation (IV-69). En revanche l'étape -3- a pratiquement disparu.

. Indépendamment de celà, les deux algorithmes jouissent des mêmes propriétés sur le plan théorique. Sur le plan numérique, l'algorithme 6 nécessite davantage de calculs (détermination de l'état pour $t = 0, \dots, N'$) auxquels il faut assurer une certaine précision particulièrement lorsque les pôles correspondants sont voisins de 1. De ce point de vue l'algorithme 5 offre une meilleure stabilité puisque tous les paramètres de la réalisation sont directement déterminés à partir des données d'entrées/sorties.

. En ce qui concerne la mise en oeuvre de ces deux derniers algorithmes, elle suit une fois de plus la technique de la factorisation orthogonale opérée au moyen des transformations de Householder. Un inconvénient de cette approche réside ici dans la taille des matrices S_{ℓ}^i . Cette difficulté doit cependant être pondérée par le fait que le problème de taille mémoire est de moins en moins crucial (mémoire virtuelle) et qu'il est toujours possible de faire explicitement appel à des mémoires auxiliaires (disques) technique à laquelle se prêtent bien les transformations de Householder. Signalons enfin qu'en présence d'échantillon très long $N > k10^3$ par exemple, on peut appliquer ces algorithmes à des données "concentrées" équivalentes obtenues par pseudo corrélation entre $\{\underline{u}_t\}$ et $\{y_t\}$ [26].

5 - REALISATION DIRECTE D'ENTREES/SORTIES QUELCONQUE (SYSTEMES CONTINUS)

Le problème essentiel réside maintenant dans la nature même des données. On peut envisager deux solutions fondamentalement différentes : la première, une approche directe, consisterait à suivre les idées développées § IV, en substituant au décalage dans le temps des données, leurs intégrations successives. Cependant l'impossibilité, évidente, de recréer l'information perdue entre deux instants consécutifs d'échantillonnage, rend assez illusoire l'ingéniosité (souvent coûteuse) dépensée dans des procédures d'intégration numérique plus ou moins sophistiquées.

La seconde voie, que nous développons dans ce paragraphe, consiste à analyser le problème en deux temps. D'abord modéliser un système discret à partir des informations échantillonnées, cf § III, puis ensuite déterminer, à partir de ce modèle discret, un modèle continu "équivalent" dans un sens à préciser.

Cependant ce que l'on vient d'affirmer comme un déroulement simple et logique demande néanmoins quelques précautions. Considérons, par exemple, le système discret suivant :

$$x_{k+1} = -\frac{1}{2} x_k + u_k \quad (\text{IV-99})$$

Il ne lui correspond aucun système continu de la forme ...

$$\dot{x} = ax + bu \quad (\text{IV-100})$$

... dont il pourrait être l'"échantillonné". Par contre, (IV-99) pourrait être obtenu à partir d'un système continu oscillant du second ordre, échantillonné à la fréquence double de sa fréquence propre, Figure 1. Mais il est évident, Figure 1, que ce n'est pas la seule solution, ce phénomène étant connu sous le nom d'"oscillations cachées".

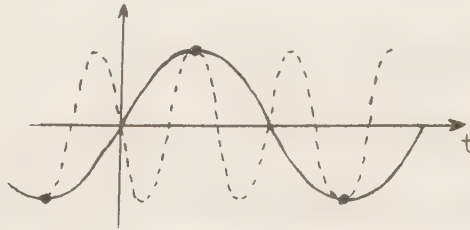


Figure 1

Nous admettons donc, comme préalable, que le système discret donné peut être effectivement considéré comme l'échantillonnage d'un certain système continu, de même dimension. Le problème d'"équivalence" se pose alors généralement en terme d'identité des sorties, aux instants d'échantillonnage, des deux systèmes excités par la même entrée.

On envisage, la plupart du temps, deux types d'entrées : l'impulsion et l'échelon. Nous avons choisi ici de développer le problème afférant à l'identité des réponses impulsionnelles, mais la solution proposée se transposerait facilement à l'entrée en échelon.

Pour des raisons qui apparaîtront un peu plus loin, il ne sera pas toujours possible d'obtenir une identité parfaite entre la séquence de pondération donnée et l'échantillonnage de la réponse impulsionnelle d'un système continu de même ordre. C'est pourquoi nous définissons l'équivalence entre système discret et continu dans les termes suivants :

Définition IV-4

Un système continu d'ordre n sera équivalent à un système discret donné, de même ordre, si, et seulement si, l'écart quadratique moyen entre l'échantillonnage de la réponse impulsionnelle de l'un et la séquence de pondération de l'autre, est minimum.

5-1. Relation d'équivalence entre système continu et discret

Considérons à cet effet le schéma suivant :

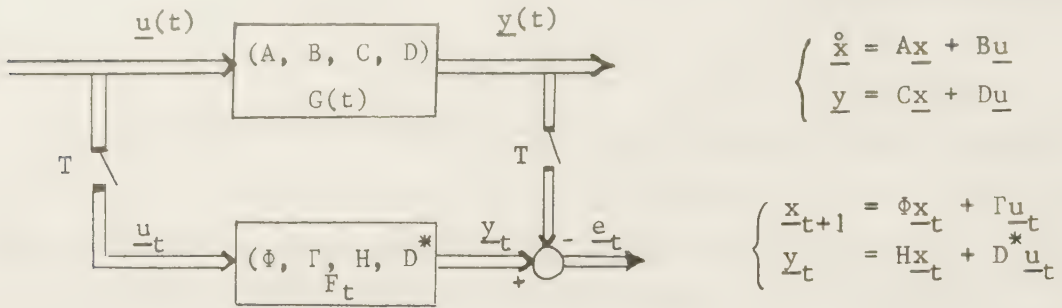


Figure 2

En se référant au Chapitre I, on a bien sûr :

$$G(t) = C e^{At} B + D\delta(t)$$

(IV-101)

$$F_t = H\Phi^{t-1}\Gamma, \quad t > 0; \quad F_0 = D^*$$

Rappelons alors rapidement certains résultats développés dans [11]

a) $\underline{u}(t)$ est un signal impulsionnel : $\underline{u}(t) \triangleq \sum_k \underline{u}_k \delta(t-kT)$

C'est un cas idéal où aucune information n'est perdue par l'échantillonnage, et pour lequel on a :

$$F_k = G(kT), \quad k > 0$$

De (IV-101) il vient immédiatement :

$$\begin{aligned} F_k = G(kT) &= C e^{AkT} B = H\Phi^{k-1}\Gamma \\ &= C(e^{AT})^{k-1} e^{AT} B \end{aligned}$$

D'où finalement :

$$\begin{cases} A = \frac{1}{T} \text{Log } \Phi \\ B = \Phi^{-1}\Gamma \\ C = H \end{cases} \quad (\text{IV-102})$$

Considérons maintenant l'instant particulier $t = 0$, et admettons que l'échantillonnage d'une impulsion soit égal à son poids, ici D . Deux hypothèses sont possibles, soit on admet que :

$$F_0 = G(0^+) + D$$

Alors de (IV-101), il vient :

$$\begin{aligned} D^* &= CB + D \\ \implies D &= D^* - H\Phi^{-1}\Gamma \end{aligned} \quad (IV-103)$$

Soit, plus vraisemblablement, on connaît D indépendamment, par exemple par le saut, à l'origine, de la réponse indicielle. Ceci nous conduit à :

$$\begin{aligned} F_0 &= G(0^+) \\ \implies D^* &= CB \\ D^* &= H\Phi^{-1}\Gamma \end{aligned} \quad (IV-104)$$

... équation qui définit des contraintes, a priori non satisfaites par les données, sauf si $\Sigma(\Phi, \Gamma, H, D^*)$ a été identifié, en tenant compte de (IV-104).

b) $\underline{u}(t)$: signal en créneaux échantillonné en début de période

Dans ce cas, réponse indicielle discrète et échantillonnage de la réponse indicielle du système continu sont identiques [11], c'est-à-dire :

$$\int_0^{kT} G(\tau) d\tau = \sum_{t=0}^k F_t$$

Nous signalons cette hypothèse de travail, plus particulièrement pour l'intérêt qu'elle présente dans le cas où l'équivalence des systèmes discrets et continus se juge sur celle des réponses indicielles, au sens de la définition (IV-4).

c) $\underline{u}(t)$ est un signal quelconque

On supposera, ici, que le pas d'échantillonnage est relativement petit par rapport à la rapidité de variation du signal d'entrée (conditions de Shannon), et qu'il n'y ait pas de corrélation entre l'instant d'échantillonnage et le signal.

Dans ces conditions, on a l'approximation : [11]

$$F_k = \int_{kT-T/2}^{kT+T/2} g(\tau) d\tau \quad (IV-106)$$

que l'on peut encore transformer en :

$$F_k = \begin{cases} \frac{T}{2} G(0^+) + D & ; k = 0 \\ T G(kT) & ; k > 0 \end{cases}$$

nouvelle approximation dont l'interprétation géométrique est illustrée à la figure 3.

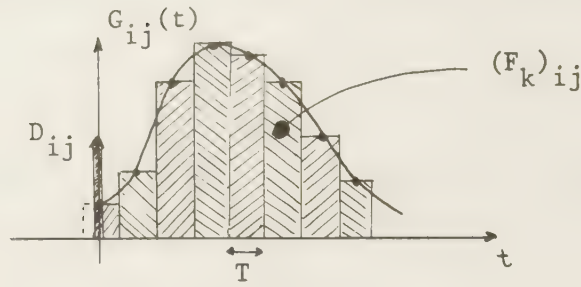


Figure 3

De (IV-107) découlent les relations suivantes :

$$\begin{aligned} \underline{k > 0} \quad H\Phi^{k-1} &= TC(e^{AT})^{k-1} e^{AT} B \\ \Rightarrow \begin{cases} A = \frac{1}{T} \text{Log } \Phi \\ B = \frac{1}{T} \Phi^{-1} \Gamma \\ C = H \end{cases} & \quad (IV-108) \end{aligned}$$

Pour $k = 0$, il vient :

$$D^* = \frac{T}{2} CB + D$$

d'où :

$$D = D^* - \frac{1}{2} H\Phi^{-1} \Gamma \quad (IV-109)$$

Compte tenu des développements du § IV, c'est ce dernier cas que nous considérerons par la suite. Résumons donc les formules de passage que nous utiliserons, soit :

$$\begin{cases} A = \frac{1}{T} \text{Log } \Phi \\ B = \frac{1}{T} \Phi^{-1} \Gamma \\ C = H \\ D = D^* - \frac{1}{2} H\Phi^{-1} \Gamma \end{cases} \quad (IV-110)$$

avec les hypothèses Φ régulières pour $n > 2$ et $\Phi = \Phi_{11} > 0$ dans le cas particulier $n = 1$. Rappelons qu'il ne peut s'agir que d'une approximation cf (IV-106). Indépendamment de ceci, il peut être intéressant d'exploiter des connaissances a priori, ou indispensable de tenir compte d'informations concernant la forme à l'instant initial ($t=0$) de certaines réponses impulsionnelles du système continu à identifier. Par exemple, pas d'impulsion : $D_{ij} = 0$ ou/et réponse

nulle à l'origine $(CB)_{kl} = 0$.

Si ces contraintes n'ont pas été prises en compte lors de la modélisation discrète, c'est-à-dire respectivement $(D^* - \frac{1}{2} H\Phi^{-1}\Gamma)_{ij} = 0$ ou/et $(H\Phi^{-1}\Gamma)_{kl} = 0$, elles pourront toujours être imposées a posteriori, mais au détriment d'une approximation sur la suite de la réponse.

La solution que l'on propose, consiste à essayer de minimiser l'écart correspondant.

IV-2. Détermination du modèle continu dans la forme canonique observable

Le passage d'un modèle discret à un modèle continu a déjà été souvent traité dans la littérature, en se basant principalement sur la transformation d'une fonction de transfert en z en une fonction de transfert en s , [27] par exemple. Dans ce domaine le plus grand nombre de suffrages vont généralement à la transformation bilinéaire bien connue $s = \frac{2}{T} \left(\frac{1-z^{-1}}{1+z^{-1}} \right)$. Lorsqu'on raisonne directement en termes d'équations d'états, les techniques préconisées passent invariablement par le calcul explicite du logarithme d'une matrice [28], [29] dont les difficultés numériques sont bien connues.

La solution, que nous allons proposer, semble nouvelle à deux titres : d'une part, elle ne requiert pas le calcul explicite du logarithme d'une matrice, d'autre part, le système continu équivalent est obtenu dans la structure canonique observable. Pour fixer les idées, nous supposerons qu'il s'agit de la forme (2) (Bucy) mais cette hypothèse n'est nullement limitative.

Le problème posé est donc le suivant : "étant donné une réalisation $\Sigma(\Phi, \Gamma, H, D^*)$ dans une forme canonique observable, déterminer le système continu équivalent ayant la même structure canonique".

La solution en est maintenant développée à travers l'...

IV-2-1. Algorithme 1

Les données sont (Φ, Γ, H, D^*) , T le pas d'échantillonnage, N caractérisant l'horizon $N.T$ sur lequel porte l'équivalence, et les couples $(i,j)-(k,l)$ relatifs aux contraintes.

- 1 - Déterminer la suite $\{F_0, F_1, \dots, F_N\}$
- 2 - Calculer les pôles du système discret, en déterminant les zéros des p polynômes caractéristiques de degré n_i , relatifs aux blocs diagonaux de Φ (matrices compagnons).

3 - En déduire les pôles du système continu au moyen de la relation :

$$\text{pôle du système continu} = \frac{1}{T} \text{Log (pôle du système discret)}.$$

4 - Construire la matrice A, bloc diagonale compagnon, de structure identique à Φ (bloc de dimensions $n_i \times n_i$, $i = 1, \dots, p$) associés aux pôles calculés en -3-

5 - Calculer la matrice $P = e^{AT}$

6 - Calculer les matrices B et C telles que : $CP^k B = \frac{1}{T} F_k$, $k = 1, 2, \dots, N$
la matrice B étant pleine et la matrice C ayant la structure suivante :

$$C = \left[\begin{array}{c|c|c|c} \text{///} & \text{///} & \text{///} & \text{///} \\ \hline \text{c}_i & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline \end{array} \right]$$

éléments à déterminer } $\xrightarrow{s_{i-1} \quad n_i \quad n}$ ligne i avec $s_i = \sum_{j=1}^i n_j$; $s_0 = 0$

7 - S'il y a des contraintes, aller en -8-. Sinon, calculer $D = F_0 - \frac{T}{2} CB$ et aller en -11-

8 - Effectuer le changement de base $S^* = \prod_{i=2}^p S_i^*$

$$S_i^* = \left[\begin{array}{c|c} \mathbf{1} & 0 \\ \hline -\frac{1}{T} & \mathbf{1} \\ \hline -\underline{c}_i & 0 \end{array} \right] \xrightarrow{S_i^* (\underline{c}_i)} C S^* = C^* = \left[\begin{array}{c|c|c|c} 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \end{array} \right]$$

ligne i

puis en déduire : $P^* = S^{*-1} P S^*$

Remarque : Si $n_i = 0$ on pose $S_i^* = \mathbf{1}_{n,n}$

$$(S_i^*)^{-1} = S_i^* (-\underline{c}_i)$$

IV-2. Démonstration

En ce qui concerne les points 1 à 5, 7 et 10, elle découle directement des équations (IV-110). Quant au point -9- il ne fait que traduire la définition IV-4.

Revenons maintenant au point -6-. Outre le fait qu'on se soit placé a priori dans une base telle que les blocs diagonaux de A soient du type compagnon, on a supposé implicitement que (Φ, Γ, H, D^*) et donc (A, B, C, D) pouvaient être exprimés dans la forme canonique découplée. Comme nous l'avons vu au Chapitre III cette hypothèse n'est pas restrictive en pratique, par suite il existe bien deux matrices C et B solution du point -6-. L'étape -8- est triviale à vérifier et a pour objet de rendre linéaire les contraintes $(CB)_{ij} = 0$. La dernière étape relève d'un esprit voisin de celui de la technique de découplage développée Chapitre III § 5, à la différence que la solution est ici explicite.

Raisonnons dans l'hypothèse où $p = 2$ et définissons une matrice S de la forme :

$$S = \left[\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline \mathbf{S}_1 & \mathbf{1} \end{array} \right] \begin{array}{l} \updownarrow n_1 \\ \updownarrow n_2 \end{array} \iff S^{-1} = \left[\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline -\mathbf{S}_1 & \mathbf{1} \end{array} \right]$$

Il vient :

$$S^{-1}AS = \left[\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline -\mathbf{S}_1 & \mathbf{1} \end{array} \right] \cdot \left[\begin{array}{c|c} \mathbf{A}_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_2 \end{array} \right] \cdot \left[\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline \mathbf{S}_1 & \mathbf{1} \end{array} \right] = \left[\begin{array}{c|c} \mathbf{A}_1 & \mathbf{0} \\ \hline -\mathbf{S}_1\mathbf{A}_1 + \mathbf{A}_2\mathbf{S}_1 & \mathbf{A}_2 \end{array} \right] \quad (\text{IV-112})$$

Il s'agit de déterminer S telle que d'une part :

$$-\mathbf{S}_1\mathbf{A}_1 + \mathbf{A}_2\mathbf{S}_1 = \left[\begin{array}{c|c} \mathbf{0} & \\ \hline \alpha_{21} & \end{array} \right] \begin{array}{l} \updownarrow n_2 \\ \updownarrow n_1 \end{array} \quad (\text{IV-113})$$

et telle que :

$$\left[\begin{array}{c|c} \mathbf{1} \mathbf{0} \dots \mathbf{0} & \mathbf{0} \dots \mathbf{0} \\ \hline \mathbf{h}^T & \mathbf{1} \mathbf{0} \dots \mathbf{0} \end{array} \right] \begin{array}{l} \updownarrow n_1 \\ \updownarrow n_2 \end{array} \star S = \left[\begin{array}{c|c} \mathbf{1} \mathbf{0} \dots \mathbf{0} & \mathbf{0} \dots \mathbf{0} \\ \hline \mathbf{0} \dots \mathbf{0} & \mathbf{1} \mathbf{0} \dots \mathbf{0} \end{array} \right] \begin{array}{l} \updownarrow n_1 \\ \updownarrow n_2 \end{array} \quad (\text{IV-114})$$

Notons :

$$S_1 = \left[\begin{array}{c|c|c|c} \underline{c}_1 & \underline{c}_2 & \cdots & \underline{c}_{n_1} \end{array} \right] = \left[\begin{array}{c} \underline{\ell}_1^T \\ \vdots \\ \underline{\ell}_{n_2}^T \end{array} \right] \quad (IV-115)$$

où $\underline{c}_i$ et $\underline{\ell}_j$ désignent respectivement les colonnes et les lignes de S_1 .

Explicitons alors (IV-114), il vient :

$$\underline{\ell}_1 = - \underline{h} \quad (IV-116)$$

puis de (IV-115), en posant :

$$A_1 = \left[\begin{array}{cccc} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ a_{11} & \cdots & & a_{1n_1} \end{array} \right] \quad \text{et} \quad A_2 = \left[\begin{array}{cccc} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ a_{21} & \cdots & & a_{2n_2} \end{array} \right]$$

On obtient :

$$- \left[\begin{array}{c|c|c|c} a_{11} \underline{c}_{n_1} & \underline{c}_1 & + a_{12} \underline{c}_{n_1} & \cdots \\ \hline \underline{c}_{n_1-1} & + a_{1n_1} \underline{c}_{n_1} & & \end{array} \right] + \left[\begin{array}{c} \underline{\ell}_2^T \\ \vdots \\ \underline{\ell}_{n_2}^T \\ \hline \sum_i a_{2i} \underline{\ell}_i^T \end{array} \right] = \left[\begin{array}{c} 0 \\ \vdots \\ \underline{\alpha}_{21}^T \end{array} \right] \quad (IV-117)$$

Il est clair que $\underline{\ell}_1$ étant connu par (IV-116), il en est de même de la première ligne de $-S_1 A_1$ d'où $\underline{\ell}_2$, d'après la première ligne de (IV-116). Connaissant $\underline{\ell}_2$ on en déduit $\underline{\ell}_3$, 2ème ligne de (IV-117), etc... Et, de la dernière ligne de (IV-117) découle les valeurs du couplage $\underline{\alpha}_{21}$ qui prend place dans A.

Dans le cas général où p est supérieur à 2, chaque matrice $\hat{S}_i$ peut être déterminée dans des conditions identiques ainsi que les couplages $\underline{\alpha}_{ij}$. Le calcul de S^{-1} par produit de S_i^{-1} n'est seulement nécessaire qu'à la construction de la nouvelle matrice B.

V-2-3. Mise en oeuvre

Reprenons successivement les différentes étapes de l'algorithme. Le Calcul de la séquence de pondération ne présente pas de difficultés particulières. En ce qui concerne le calcul des zéros d'un polynome, on dispose actuellement

d'excellentes procédures [12], [13]. La détermination de A est alors triviale puisqu'il s'agit de reconstituer des polynômes dont on connaît les zéros. Nous insisterons davantage sur le calcul de P dont la précision conditionne la signification de tous les calculs qui suivent. Une procédure qui nous est apparue particulièrement satisfaisante s'inspire du travail de Kallstrom [14], et consiste à utiliser un développement en série, appliqué non pas directement à A, mais à une matrice déduite de celle-ci par une mise à l'échelle appropriée. Les termes de la série sont alors calculés en tenant compte de la forme compagnon des blocs diagonaux de A, puis sommés d'une manière optimale [15], vis à vis de la précision machine.

A l'étape suivante, point - 6 -, le calcul des matrices B et C est extrêmement simple puisqu'il se réduit à résoudre des systèmes d'équations linéaires. En effet, considérons (pour simplifier l'exposé) la disposition des opérations dans le cas où $p = 2$.

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & | & 0 & \dots & 0 \\ \hline \underline{c}_2^T & & & & & | & 1 & 0 & \dots & 0 \end{bmatrix} \times \begin{bmatrix} P_1 & 0 \\ \hline 0 & P_2 \end{bmatrix}^k \times \begin{bmatrix} B_1 \\ \hline B_2 \end{bmatrix} \begin{matrix} \updownarrow n_1 \\ \updownarrow n_2 \end{matrix} = \frac{1}{T} F_k \Delta = \frac{1}{T} \begin{bmatrix} (f_k^1)^T \\ \hline (f_k^2)^T \end{bmatrix} \begin{matrix} \updownarrow m \\ \updownarrow 2 \end{matrix}$$

(IV-118)

Considérons la première ligne de l'équation (IV-118), il vient :

$$1^{\text{ère}} \text{ ligne de } (P_1)^k \times B_1 = \frac{1}{T} (f_k^1)^T, \quad k = 1, 2, \dots, n_1 \quad (\text{IV-119})$$

Par conséquent la détermination de B_1 à partir de (IV-119) consiste à résoudre un système linéaire à n_1 équations, n_1 inconnues et m seconds-nombres.

Maintenant, connaissant B_1 , explicitons la deuxième ligne de l'équation (IV-118) :

$$\underline{c}_2^T (P_1)^k B_1 + [1^{\text{ère}} \text{ ligne de } (P_2)^k] B_2 = \frac{1}{T} (f_k^2)^T$$

d'où

$$[1^{\text{ère}} \text{ ligne de } (P_2)^k] B_2 = \frac{1}{T} (f_k^2)^T - \underline{c}_2^T (P_1)^k B_1 \quad (\text{IV-120})$$

$$k = 1, 2, \dots, n_2$$

ce qui nous ramène au cas précédent. Bien que théoriquement utiliser plus de n_1 équations pour calculer B_1 n'apporte aucune information supplémentaire, il est en pratique recommandé de résoudre des systèmes à N équations au sens des moindres carrés suivant toujours la méthode de Householder [1]. Cette technique s'étend d'ailleurs efficacement au problème avec contraintes linéaires égalités [19], solution préconisée pour calculer B^* , au point - 9 -.

6 - CONCLUSION

Nous avons successivement abordé, dans ce chapitre, les trois principaux thèmes relatifs au problème de réalisation : Réponses impulsionnelles ou séquences de pondération - fonctions de transfert en s ou en z . - entrées/sorties quelconques.

Traité dans le contexte discret, parce que le plus naturel, face à la mécanisation des algorithmes, cette approche représente néanmoins une voie intéressante pour modéliser des systèmes continus.

Nous retiendrons de cette étude les résultats essentiels suivants :

- Les algorithmes 1 à 6 reposent sur un principe unique dont la mise en oeuvre est à la fois simple et numériquement stable
 - Certains d'entre eux, algorithmes 2, 4 et 6, apparaissent pour la première fois à notre connaissance
 - L'algorithme 7 relatif au passage discret - continu, semble nouveau et présente deux caractéristiques particulièrement importantes ; il ne requiert pas le calcul du logarithme d'une matrice, le système continu équivalent est obtenu dans la même forme canonique que le système discret dont il est issu.
-
-

BIBLIOGRAPHIE

- [1] P. BUSINGER
G.H. GOLUB
"Linear least squares solutions by householder transformations"
Numer Moth. 7, p. 269-276 (1965)
- [2] B.L. HO
R.E. KALMAN
"Effective construction of linear state variable models from input-output functions"
Regelungs-technik, Heft 12, Ju 1966, pp. 545-548
- [3] R.E. KALMAN
P.L. FALB
M.A. ARBIB
"Topics in mathematical system theory"
Mc GRAW HILL Book Company 1969
- [4] C. BONIVENTO
R. GUIDORZI
"Canonical input-output description of linear multivariable systems"
Ricerche di Automatica, Vol 2 n°1,
January 1971, pp. 72-83
- [5] Ph. DE LARMINAT
A. BARRAUD
"Un algorithme de réalisation minimale à partir de données d'entrée/sortie"
C.R.A.S. Paris t. 275 (24 Juillet 1972)
Série A pp.315-318
- [6] K. GLOVER
J.C. WILLEMS
"On the identifiability of linear dynamical systems"
Congrès I.F.A.C. La Haye 1973, papier TO-2
- [7] E. TSE
J.J. ANTON
"On the identifiability of parameters"
I.E.E.E. Trans. Aut. Cont. Vol. AC 17 n° 5
October 1972, pp. 637-646
- [8] E. TSE
H. WEINERT
"Correction and extension of-on the identifia-
bility of parameters"
I.E.E.E. Trans. Aut. Cont. Vol. AC 18 n° 6
December 1973, pp. 687-682
- [9] K.J. ASTROM
P. EYKHOFF
"System identification, a survey"
I.F.A.C. Congrès 1970 PRAGUE, Survey-paper
- [10] A. BARRAUD
Ph. DE LARMINAT
"Minimal réalisation and approximation of linear systems from normal operating records"
3th I.F.A.C. symposium, La Haye, Juin 1973
paper TS-5
- [11] A. BARRAUD
"Contribution à l'identification des séquences de pondération des systèmes dynamiques"
Thèse Docteur-Ingénieur, 12 Juillet 1971 NANTES

- [12] M.A. JENKINS "Three stage variable shift iterations for the solution of polynomial equations with a posteriori error bounds"
Rep. C.S. 138 computer Sci Dept. STANFORD U. CALIFORNIA 1969
- [13] K. MADSEN "A root-finding algorithm based on Newton's method"
Bit 13 (1973) pp. 71,75
- [14] J.C. KALLSTROM "Computing e^A and $\int e^{Ac} dc$ "
Report 7309 March 1973 Lund Institute of Technology
- [15] M.A. MALCOLM "On accurate floating-point sommation"
C.A.C.M. Vol. 14 n° 11, 1971, pp. 731, 736
- [16] R.J. HANSON "Extensions and applications of the householder algorithm for solving linear least squares problems"
C.L. LAWSON
Math. of Comp. October 1969, Vol. 23, n° 108 pp. 787, 812
- [17] E.G. GILBERT "Controllability and observability in multivariable control systems"
SIAM J. Control Vol 2 n° 1, pp. 128-151 (1963)
- [18] R.E. KALMAN "Irreducible realisation and the degree of a rational matrix"
J. Soc. Ind. App. Math. Vol 13 n° 2 June 1965 pp. 520-544
- [19] H.H. ROSENBROCK "Computation of minimal representation of a rational transfer function matrix"
Proc. I.E.E., Vol. 115 n° 2, Février 1968 pp. 325-327
- [20] C.T. CHEN "A simplified irreducible realisation algorithm"
D.P. MITAL
I.E.E.E. Trans. Aut. Cont. Auguste 1972, pp. 535-537
- [21] J.E. ACKERMANN "Canonical minimal réalisation of a matrix of impulse response sequense"
R.S. BUCY
Information and Control, Vol. 19, 1971, pp. 224-231
- [22] B. GOPINATH "On the identification of linear time invariant systems from input-output data"
The Bell system tech. J. Vol 48 n° 5 1969 pp. 1101, 1113

- [23] M.A. BUDIN "Minimal realization of discrete linear systems from input-output observations"
I.E.E.E. Trans. Aut. Cont. October 1971,
pp. 395-401
- [24] J.E. ACKERMANN "Die minimale Ein Ausgangs Beschreibung von Mehrgroßen systemen und ihre Bestimmung aus ein Ausgangs Messungen".
Regelungs-technik Heft 5, Vol 19, Jahrgang 1971
pp 203-206
- [25] D.Q. MAYNE "A canonical model for identification of multi-variable linear systems"
I.E.E.E. Trans. on Aut. Cont, October 1972
pp. 728, 729
- [26] Ph. DE LARMINAT "Concentration de données, identification de systemes en boucle fermée"
Annales E.N.S.M. 1er sem. 1972
- [27] N.K. SINHA "Estimation of transfert function of continuous system from sampled data"
Proc. I.E.E. Vol. 119, n° 5 May 1972
pp. 612-614
- [28] D.H. MEC "Finding a linear continuous system from its sampled representation"
Proc. I.E.E. Vol. 118, n° 7, July 1971
pp. 943-945
- [29] W. GERTH "On the construction of low order linear state variable models from measured data"
I.F.A.C. Dusseldorf October 1971, paper 1.2.1.

Chapitre V

APPROXIMATIONS OPTIMALES ALGORITHMES - RESULTATS EXPERIMENTAUX

o o
o

- 1 - Introduction
 - 2 - Approximation des systèmes monovariabiles
 - 3 - Résultats expérimentaux (systèmes monovariabiles)
 - 4 - Approximation des systèmes multivariabiles
 - 5 - Résultats expérimentaux (systèmes multivariabiles)
 - 6 - Commentaire sur les méthodes numériques
 - 7 - Conclusion
- Bibliographie

Résumé :

La modélisation d'un système physique est par nature même un problème d'approximation. Celui-ci est essentiellement abordé d'un point de vue déterministe, complété le cas échéant par quelques remarques issues d'hypothèses probabilistes.

Nous traitons, en détail, le cas des systèmes monovariabiles qui autorise des comparaisons cohérentes et représente un support utile dans le domaine plus général des systèmes multivariabiles.

On présente à cet effet un ensemble de résultats expérimentaux, effectués en simulations, et mettant en évidence les points essentiels de ces développements.

Les difficultés propres aux systèmes multi-sorties nous ont cependant incité à emprunter, quelquefois, des voies plus pragmatiques en se laissant guider par le souci de conserver, aux solutions, un rapport coût-calcul/performance-significatif.

Les nombreux tests numériques, effectués tant en simulations que sur processus réels, sont là pour démontrer le bien fondé d'une telle approche.

1 - INTRODUCTION -

Nous avons adopté jusqu'ici certaines hypothèses de travail, relatives aux systèmes dynamiques considérés : linéarité, invariance, paramètres localisés. Sur cette base, a été développé un ensemble d'algorithmes de réalisation conduisant à des modèles ayant une structure canonique irréductible.

Cependant lorsqu'on est confronté avec un système physique, il n'est généralement pas possible de définir, dans des termes aussi précis, les caractéristiques dynamiques réelles du processus et de son environnement. En effet ce système peut présenter, par exemple, une dérive plus ou moins lente, ne pas être rigoureusement linéaire, subir des perturbations, etc... Tous ces aspects contribuent à donner aux problèmes de réalisation un caractère d'approximation qu'il n'avait pas été nécessaire d'envisager jusqu'alors, et le plus souvent, beaucoup plus difficile à résoudre.

On serait tenté, dans ces conditions, d'effectuer des développements plus sophistiqués, susceptibles d'offrir un plus large domaine d'applications. La modélisation stochastique de l'environnement (perturbations, bruits...) en est un exemple. Cependant, là encore, l'aboutissement d'une telle approche reste subordonné à d'autres hypothèses : variables gaussiennes, spectre rationnel, etc... qui définissent, en soi, de nouvelles limites.

Notre propos n'est pas d'approfondir de telles directions, mais plutôt de conserver les hypothèses initiales et de se laisser guider, quant au choix d'un critère d'approximation, par l'interprétation qu'on pourrait lui attribuer sous certaines hypothèses et par la signification qu'il peut avoir d'un point de vue physique ou pratique.

Par ailleurs, on fera remarquer que cette notion d'approximation, qui apparaît maintenant à un niveau conceptuel (des algorithmes) intervenaient en fait déjà implicitement dans la mise en oeuvre des algorithmes de réalisation, en raison de la précision finie des calculs. Rappelons, en effet, que les procédures 1 à 6 (cf Ch. IV, § 4) reposent sur la dépendance, ou non dépendance linéaire de certains vecteurs par rapport à d'autres. L'utilisation des transformations de Householder, telle qu'elle a été faite, revient à admettre qu'un vecteur donné est linéairement indépendant d'un ensemble d'autres vecteurs lorsque celui-ci est distant (en valeur relative) de l'espace engendré par ces derniers, d'une quantité supérieure à la précision machine.

Néanmoins l'application directe de ces procédures à l'ajustement inconditionnel d'un modèle aux données d'un processus physique conduit la plupart du temps à des dimensions prohibitives, inexploitablement directement, et motivant une

étape supplémentaire: la réduction du modèle. Les techniques d'approximation que nous allons développer dans ce chapitre représentent une autre approche dont l'objectif essentiel est d'éviter cette modélisation intermédiaire.

Une première solution (qui vient immédiatement à l'esprit) consisterait à prendre comme critère d'approximation, précisément le degré d'indépendance (des vecteurs dont on vient de parler plus haut) que l'on exprimerait en substituant à la précision machine une quantité plus au moins petite.

D'autres méthodes sont bien sûr envisageables, en raisonnant par exemple sur les caractéristiques internes du modèle (observabilité-accessibilité) ou au contraire sur son comportement externe (erreur entre sortie modèle et sortie processus). Dans le cadre de cette dernière approche on présente une application originale de certains résultats récents d'optimisation non linéaire dont l'efficacité a été abondamment testée § 13.

Le problème multivariable, pour lequel les résultats obtenus dans ce domaine sont encore relativement limités, présente, quant à lui, des difficultés spécifiques certaines, notamment en ce qui concerne la détermination d'une structure "optimale" § 4. C'est ce point qui rend précisément difficile la transposition des résultats, acquis pour les systèmes monovariables, aux systèmes multi-sorties. L'expérience montre, d'ailleurs, qu'un traitement réaliste de ces systèmes ne peut être que le fruit d'un compromis, entre un certain degré de sous-optimalité dans la solution, et la complexité du problème numérique correspondant à résoudre. Précisons, enfin, que nous nous placerons ici uniquement dans un cas discret en raison de ce qui a été dit au chapitre IV (§ 5) à propos du passage d'un modèle discret au modèle continu "équivalent".

2 - APPROXIMATION DES SYSTEMES MONOVARIABLES -

L'ensemble des techniques qui se rattachent à ce thème, tant dans le cadre des systèmes continus que discrets, relève essentiellement de quatre idées maîtresses :

- élimination de certains modes (associés aux petites constantes de temps)
- développement de la transmittance en fraction continue
- minimisation d'un critère dans le domaine fréquentiel (complexe : gain-phase).
- minimisation d'un critère dans le domaine temporel.

Précisons plus en détail ces différents points. La philosophie de la première méthode, qui consiste à éliminer purement et simplement les constantes de temps les plus petites, s'appuie sur le fait que leur contribution à la réponse du système devient rapidement négligeable [1]. Malheureusement cette approche n'est pas toujours bien adaptée aux objectifs de commande. Par ailleurs, malgré l'attrait de sa simplicité, une fois les pôles calculés, on peut se demander si une "meilleure" approximation ne pourrait pas être obtenue en laissant un plus grand degré de liberté aux pôles du système réduit.

C'est en fait l'objectif principal des méthodes de développement en fraction continue de la transmittance [2], technique développée initialement pour les systèmes continus par C.F. CHEN et L.S. SHIEH [3]. Elles sont intéressantes à deux titres :

- leur simplicité : modèle réduit d'ordre n obtenu à partir des $2n$ premiers termes d'un développement, dont la convergence est très rapide ;
- les relations que l'on peut imposer entre modèle original et modèle réduit en ajustant certains moments [†] de la séquence de pondération où/et certains paramètres de Markov [4]. On peut ainsi ajuster transitoires et régime permanent du système réduit. Ce résultat n'est cependant pas toujours aisé à obtenir parce que la stabilité du système original n'implique pas celle du système réduit.

On peut se demander alors pourquoi ne pas essayer d'atteindre cet objectif, assez naturel, en explicitant le problème au moyen de la minimisation d'un critère approprié. Les deux dernières approches, citées plus haut, tentent précisément de répondre à cette question. L'ajustement des courbes de fréquence dans le plan complexe étant essentiellement l'apanage des systèmes continus, nous ne mentionnerons que l'article de BOSLEY et LEES [5]^{††} qui font le point sur la question.

[†] $\{f_k\}$ étant la séquence de pondération, ses moments sont définis par $m_i = \sum_{k=0}^{\infty} f_k k^i$

^{††} Signalons que ces auteurs traitent dans son ensemble le problème de réduction des systèmes continus sur la base d'une cinquantaine de références.

Dans le domaine temporel, plusieurs voies ont été proposées suivant la définition de l'erreur à minimiser. D. MITRA [6], minimise, par exemple, un critère de performance qui mesure les mérites relatifs, au niveau de l'erreur de sortie, entre plusieurs sélections de pôles du modèle original. Une autre solution, beaucoup plus simple, et bien connue en identification, consiste à minimiser l'erreur d'équation d'un modèle réduit de degré n (moindres carrés), technique qui assure la meilleure prédiction de la sortie à l'instant $k + 1$ (erreur quadratique), en utilisant les mesures entrées/sorties de l'instant $k - n$ à l'instant k [7]. Divers auteurs redécouvrent périodiquement cette approche en lui attribuant d'ailleurs parfois [8] des propriétés erronées : minimisation de l'erreur de sortie, au sens de la norme euclidienne, objectif par ailleurs extrêmement intéressant à atteindre, en raison de sa signification précise. En fait ce dernier type de critère, minimisation d'une fonctionnelle de l'erreur de sortie, est sans doute la meilleure façon de mesurer la qualité de l'approximation faite, en l'absence d'informations précises sur l'objectif final de la modélisation du système. Il faut cependant retenir que cette optimisation ne peut être explicite, de sorte que les performances que l'on peut attendre de ce genre de méthode dépend considérablement de la façon dont est traitée la minimisation du critère. Nous allons développer, dans ce sens, une nouvelle approche faisant appel aux résultats les plus récents de la programmation non linéaire, grâce auxquels on peut prétendre à des algorithmes extrêmement robustes et incomparablement plus efficaces que certaines solutions classiques du problème [9]. Auparavant nous voudrions compléter ces considérations générales par les réflexions suivantes :

- La mesure effective de la qualité de l'approximation n'a fait appel jusqu'ici qu'à des informations externes (couples d'entrées/sorties par exemple).

- En raisonnant sur la représentation interne des systèmes, on pourrait introduire un critère exprimant un degré d'accessibilité ou/et d'observabilité qui ne pourrait dépendre que des paramètres du système.

- L'ordre du système réduit est donné, alors qu'il pourrait offrir un degré de liberté supplémentaire au problème.

A propos de cette dernière observation, imaginons un système du 4^{ème} ordre ayant deux paires de pôles complexes conjugués, il est alors impossible par exemple, de construire un modèle réduit du 3^{ème} ordre dont les pôles soient issus d'une sélection de ceux du système original. Par ailleurs, en s'imposant un ordre a priori, éventuellement trop petit, on s'expose au risque d'avoir un modèle réduit dont la dynamique n'ait plus aucune signification.

Revenons maintenant à notre propos initial, à l'occasion duquel on a introduit et utilisé jusqu'alors, presque indifféremment les termes "réduction" et "approximation". Nous voudrions, pour la suite de cet exposé, leur attribuer néanmoins une signification quelque peu différente.

On parlera d'*approximation* quand il s'agira de déterminer directement un modèle réduit à partir d'informations relatives au processus-objet, telles que des couples entrées-sorties (séquences de pondérations et réponses indicielles inclus).

Le terme *réduction* sera réservé au problème correspondant, lorsqu'on dispose déjà initialement d'un modèle (jugé trop complexe).

Si l'on se réfère aux Figures 1 et 2 qui illustrent ces définitions, il est clair que les mêmes algorithmes peuvent être utilisés dans l'un et l'autre cas mais avec des données différentes.

Ces remarques devraient expliquer l'optique adoptée au cours des développements qui suivent, où réduction et approximation sont traitées parallèlement en raison de leurs objectifs communs.

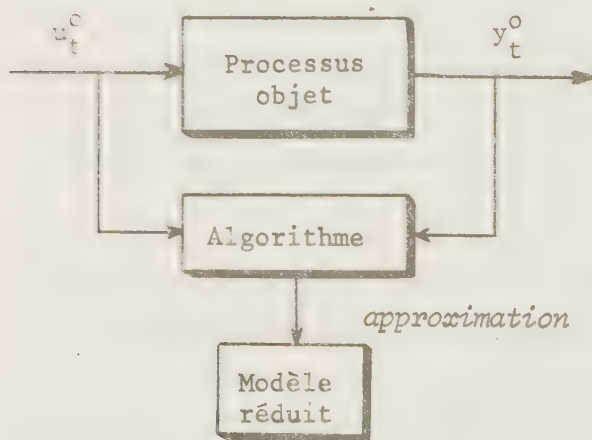


Figure 1

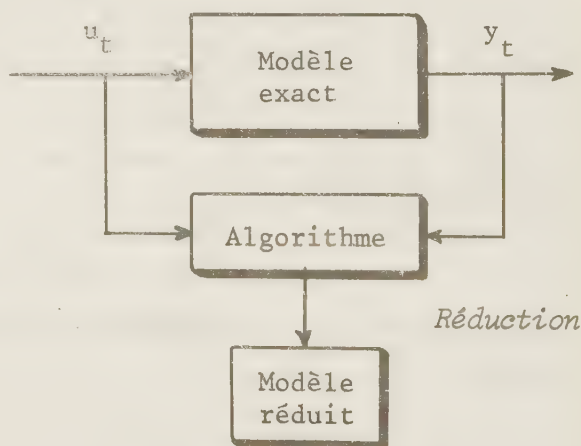


Figure 2

2.1 - Définition du critère d'approximation

Soit $\{u_t, y_t\}$ les couples d'entrées/sorties associées à un système Σ_k , et liés par la récurrence :

$$y_t = a_1 y_{t-1} + \dots + a_k y_{t-k} + b_c u_t + \dots + b_k u_{t-k} \quad (V-1)$$

que l'on peut récrire sous forme matricielle pour $t = k, \dots, N$:

$$\underline{y}_k = \tilde{R}_{k,N} \underline{a}_k \quad (V-2)$$

en notant :

$$R_{k,N} \triangleq \begin{bmatrix} u_0 & y_0 & & & u_k & y_k \\ \vdots & \vdots & & & \vdots & \vdots \\ u_{N-k} & y_{N-k} & & & u_N & y_N \end{bmatrix} \triangleq [\tilde{R}_{k,N} \mid y_k] \quad (V-3)$$

$\xleftarrow{\hspace{10em}} 2(k+1)$

et

$$\underline{a}_k^T \triangleq [b_k \quad a_k \quad \dots \quad b_1 \quad a_1 \quad b_0] \quad (V-4)$$

$\xleftarrow{\hspace{10em}} 2k+1$

Rappelons enfin que si Σ_k est identifiable à partir des couples entrées/sorties $\{u_t, y_t\}$, $\tilde{R}_{k,N}$ est de rang maximum $2k+1$, pour tout N supérieur ou égal à $4k$.

En conséquence, étant donné un modèle d'ordre n , identifiable à partir des courbes entrées/sorties $\{u_t, y_t\}$, il existe un certain $\underline{a}_n$ tel que :

$$\underline{y}_n = \tilde{R}_{n,N} \underline{a}_n \quad (V-5)$$

mais :

$$\underline{y}_k \neq \tilde{R}_{k,N} \underline{a}_k \quad \forall \underline{a}_k, k < n \quad (V-6)$$

Posé dans ces termes, trouver un $\underline{a}_k$ optimisant un certain critère de qualité est un problème de *réduction*.

Dans un contexte différent la relation (V-6) caractérise également le problème d'approximation lorsque les couples $\{u_t, y_t\}$ sont des mesures effectuées sur un processus en fonctionnement, avec toutefois la différence qu'il n'existe généralement pas de valeur pour n ($< N/4$, pour l'identifiabilité) telle que (V-5) soit satisfaite.

Par suite les deux problèmes peuvent être confondus pour leur traitement dans le cadre de critères s'exprimant en termes d'entrées-sorties.

2.1.1 - Critère d'erreur d'équation

Soit $\underline{e}_n$ le vecteur d'erreur, de composantes $e_t, t = n, \dots, N$, tel que :

$$\underline{y}_n = \tilde{R}_{n,N} \underline{a}_n + \underline{e}_n \quad (V-7)$$

Définition 3.1 - Le scalaire e_t est ce qu'on appelle l'erreur d'équation du modèle Σ_n , auquel on associe le critère :

$$J' \triangleq ||\underline{e}_n||^2 = \sum_{t=n}^N e_t^2 \quad (V-8)$$

Bien que cette erreur n'ait pas une signification physique directe (cf. Fig. 3), elle présente néanmoins deux caractéristiques très intéressantes.

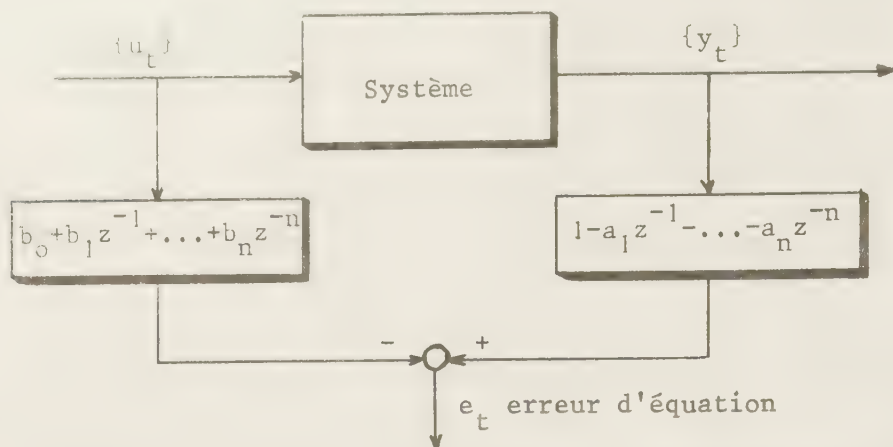


Figure 3

On notera d'abord que la minimisation de J' par rapport à $\underline{\alpha}_n$ est en principe aisée, puisque J' est quadratique ; on est alors ramené au problème classique de moindres carrés dont la solution unique $\underline{\alpha}_n^*$ ($\hat{R}_{n,N}$ étant nécessairement de rang maximum) peut être efficacement obtenue en résolvant directement (V-7) par la méthode de Householder [10].

Par ailleurs, notons y_t^* la sortie prédite à l'instant t , et définie par :

$$y_t^* = a_1^* y_{t-1} + \dots + a_n^* y_{t-n} + b_0^* u_t + \dots + b_n^* u_{t-n} \quad (V-9)$$

Il est clair que :

$$e_t = y_t - y_t^* \quad (V-10)$$

Relation dont découle la ...

Propriété 5.1 Le modèle $\Sigma_n(\underline{\alpha}_n^*)$ obtenu par minimisation de l'erreur d'équation est celui qui assure la meilleure prédiction (au sens des moindres carrés) de la sortie à l'instant t , connaissant les entrées/sorties passées (u_k, y_k) , $k = t-1, \dots, t-n$.

Remarque : Cette propriété n'est vraie que pour le couple d'entrées/sorties à partir duquel les paramètres $\underline{\alpha}_n^*$ ont été calculés. Cependant à la limite, cette propriété est encore vérifiée pour des entrées/sorties ayant les mêmes caractéristiques statistiques.

2.1.2 - Critère d'erreur de sortie

On considère de nouveau le modèle $\Sigma_n(\underline{\alpha}_n)$ auquel on applique la séquence d'entrée $\{u_t\}$. Soit $\{\eta_t\}$ la sortie modèle correspondante. D'après (V-1),

on a bien sûr :

$$\eta_t = a_1 \eta_{t-1} + \dots + a_n \eta_{t-n} + b_0 u_t + \dots + b_n u_{t-n} \quad (V-11)$$

Définition 5.2 - L'erreur de sortie relative au modèle Σ_n est la quantité :

$$\epsilon_t = y_t - \eta_t \quad (V-12)$$

à laquelle on associe le critère :

$$J = \left\| \underline{\epsilon}_n \right\| \Delta \sum_{t=n}^N \epsilon_t^2 \quad (V-13)$$

La Fig. 4 illustre clairement cette seconde approche dont l'interprétation se passe cette fois de commentaires.

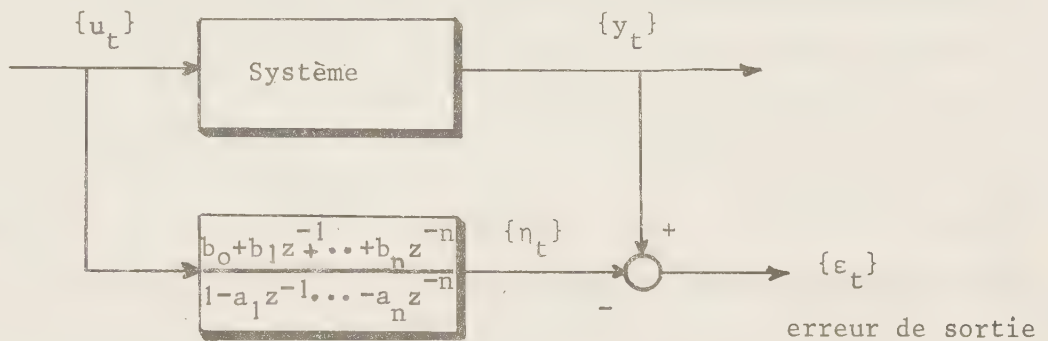


Figure 4

Revenons maintenant à la définition ϵ_t , on notera que, contrairement à (V-9), la récurrence (V-11) n'est pas uniquement définie à partir des données $\{u_t, y_t\}$; en particulier pour $t = n$, le calcul de η_n nécessite la connaissance de conditions initiales $\underline{\eta}_0^T = (\eta_0, \eta_1, \dots, \eta_{n-1})$.

De cette remarque en découle une seconde concernant l'optimisation de J qui devra, de toute évidence, être faite maintenant à la fois par rapport à $\underline{\alpha}_n$ et $\underline{\eta}_0$. Cependant lorsque l'horizon N est largement supérieur au temps de réponse du système (ce qui suppose ici la stabilité asymptotique) la contribution de $\underline{\eta}_0$ dans l'optimisation de J est souvent négligeable. Dans ce cas, on posera avantageusement $(\underline{\eta}_0)^T = (y_0, y_1, \dots, y_{n-1})$, la minimisation de J étant alors uniquement effectuée par rapport à $\underline{\alpha}_n$.

Un dernier aspect du problème porte sur la minimisation proprement dite de J . C'est à ce niveau qu'apparaît la différence fondamentale entre ces deux approches, conséquence de la nature respective des fonctions $e_t(\underline{\alpha}_n)$ et $\epsilon_t(\underline{\alpha}_n, \underline{\eta}_0)$. En effet, au vue de (V-9), e_t est une fonction linéaire des paramètres $\underline{\alpha}_n$ (J quadratique), ce qui assure la simplicité remarquable de cette première méthode. Au contraire, en ce qui concerne ϵ_t , on constatera facilement

(en écrivant par exemple (V-11) pour $t = n$ et $t = n+1$) que si l'erreur de sortie est encore une fonction linéaire de $b_0, b_1, \dots, b_n$ et η_0 , ϵ_t est non linéaire par rapport à $a_1, a_2, \dots, a_n$. La minimisation de J ne peut être alors explicitée et requiert, par suite, l'utilisation des techniques de programmation non linéaire (cf § 2.3).

Cette approche n'en n'offre pas, pour autant, moins d'intérêt, en raison des propriétés spécifiques qu'on peut lui attribuer dans un contexte probabiliste approprié. Rappelons simplement que si ϵ_t est assimilable à une variable aléatoire indépendante gaussienne, alors la solution $\hat{\underline{a}}_n$ minimisant J , est aussi la meilleure estimation possible des paramètres du système, au sens du maximum de vraisemblance. En termes statistiques $\hat{\underline{a}}_n$ définit un ensemble de variables gaussiennes dont asymptotiquement l'espérance mathématique est égale à la vraie valeur des paramètres et dont la variance est minimale. Quoiqu'il en soit ce traitement des informations $\{u_t, y_t\}$ est souvent considéré, et à juste titre, comme une solution de référence, en raison de l'optimalité de cette approche dans l'hypothèse gaussienne.

2.2 - Minimisation de l'erreur de sortie

La minimisation du critère J , défini en (V-13) doit, dans le cas général, être réalisée par rapport à $\underline{a}_n$ et $\underline{\eta}_0$, c'est-à-dire $3n + 1$ paramètres. Pour des systèmes d'ordre usuel (3 à 6 par exemple), cela représente un problème non linéaire d'environ 10 à 20 variables, donc de relativement petite taille. Cependant la très grande sensibilité du critère par rapport aux a_i rend assez souvent l'optimisation assez délicate. Les méthodes heuristiques traditionnelles telles que HOOKE et JEEVES [11] ROSENBROCK [12] ou NELDER-MEAD [13] sont alors à déconseiller en raison de leur progression extrêmement lente. Par ailleurs le calcul analytique du gradient étant assez simple, l'emploi de méthodes plus sophistiquées ne présente pas de difficultés. Notre expérience dans ce domaine nous amènera à préconiser les algorithmes du type "quasi-Newton" [14] dont la mise en oeuvre proposée par FLETCHER [15] s'est révélée des plus satisfaisantes.

Cependant, dans le cas présent où le critère se présente sous la forme d'une somme de carrés, il peut être intéressant (principalement lorsque l'optimum J est voisin de zéro de faire appel à des algorithmes de régression non linéaire. L'approche la plus connue est sans doute la méthode de Marquardt [16] dont s'inspire le programme de FLETCHER [17] qui s'est montré particulièrement efficace dans ce domaine.

Quoiqu'il en soit, indépendamment de toute technique d'optimisation, la dimension de l'espace dans lequel la minimisation de J doit être effectuée peut être *considérablement* réduite si l'on tient compte de la linéarité de ε_t par rapport aux $2n + 1$ variables $b_0, \dots, b_n$ et η_0 , en définissant un *critère projeté* $\tilde{J}$ ne dépendant plus que des n variables non linéaires $a_1, \dots, a_n$.

2.2.1 - La méthode du critère projeté

Cette méthode trouve ses origines dans les problèmes de régression non linéaire ou d'ajustement de courbe. Son objectif est de réduire la dimension de l'espace dans lequel l'optimisation doit être réalisée en exploitant la structure du critère. Bien que l'idée sur laquelle repose cette technique soit déjà connue et utilisée depuis quelques années (dans le domaine où nous l'avons situé), la démonstration du procédé en est cependant récente [18]. Nous pensons, par ailleurs, que cette approche est appliquée, ici, pour la première fois dans le cadre de l'identification des systèmes (approximation-réduction) domaine où elle est a priori séduisante dans la mesure où la dimension du problème non linéaire est réduit à l'ordre du système quel que soit le nombre d'entrées.

Pour en exposer le principe, nous considérons le problème suivant qui consiste à minimiser la norme euclidienne d'un vecteur "résidu" $\underline{r}$ de la forme :

$$\underline{r} = \underline{y} - A(\underline{\omega})\underline{v} \quad (V-14)$$

où $\underline{v}^T \Delta(v_1, \dots, v_n)$ désigne l'ensemble des paramètres qui interviennent linéairement dans le résidu $\underline{r}$ et $\underline{\omega}$ les p autres paramètres par rapport auxquels $\|\underline{r}\|$ doit être également minimisé. Notons m le nombre de composantes de $\underline{r}$ et $\underline{y}$ supposé supérieur (ou égal.) à $(n + p)$ le nombre total de paramètres. Dans ces conditions A est une matrice $m \times n$ dont les éléments sont des fonctions quelconques des ω_i . Nous admettons, ici, que cette matrice est de rang maximum (donc n), hypothèse correspondant au problème d'approximation considéré plus haut.

L'esprit de la méthode est la suivante : la minimisation d'un critère quadratique étant explicite (projection), il suffit de ne considérer que les points de l'espace R^{n+p} où le critère $I \Delta \|\underline{r}\|_2^2$ est minimum par rapport à $\underline{v}$. En d'autres termes, pour un $\underline{\omega}$ donné on ne calcule que la valeur du critère I associée à la seule valeur de $\underline{v}$ solution du sous problème d'optimisation :

$$\min_{\underline{v}} I(\underline{v}, \underline{\omega}) \Delta \tilde{I}(\underline{\omega}) \quad (V-15)$$

Il est alors démontré que [19] :

$$\min_{\underline{\omega}} \tilde{I}(\underline{\omega}) = \min_{\underline{v}, \underline{\omega}} I(\underline{v}, \underline{\omega}) \quad (V-16)$$

Cette hiérarchisation du problème d'optimisation est donc extrêmement efficace puisqu'elle résout la minimisation de I , dans R^{n+p} par une *optimisation non linéaire par rapport aux seules p variables $\underline{\omega}$* .

A titre d'illustration, on a considéré, à la figure 5, un problème à deux variables ($p = n = 1$). En effectuant la minimisation de $\tilde{I}$, on évolue nécessairement sur la trajectoire des minima de I par rapport à v de sorte qu'on atteigne effectivement l'optimum en appliquant une technique de programmation non linéaire à ω seul, (on en a fait figurer 5 itérations hypothétiques en exemple).

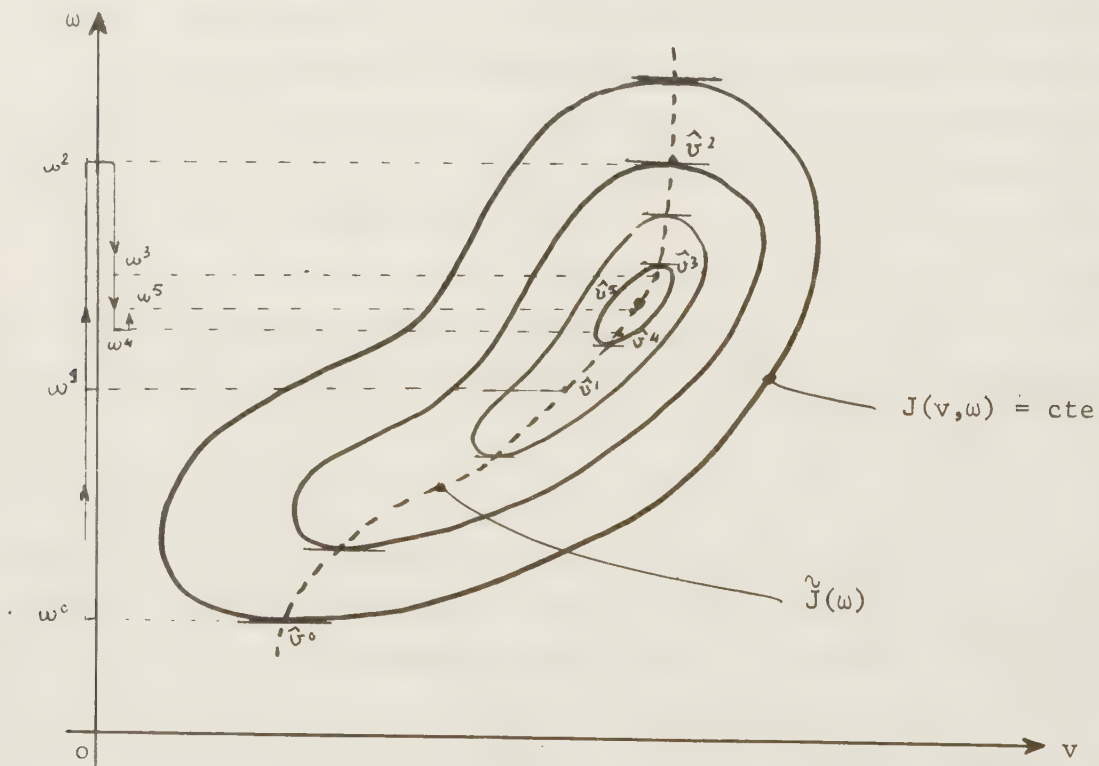


Figure 5

2.2.2 - Gradients et Jacobien associés à I et $\tilde{I}$

Les méthodes de programmation non linéaire auxquelles on a précédemment fait référence font appel au calcul du gradient du critère ou du Jacobien des résidus. Il convient donc de préciser ces quantités.

De (V-14), il vient :

$$\text{Jacobien de } \underline{r} \left\{ \begin{array}{l} \frac{\partial \underline{r}(\underline{v}, \underline{\omega})}{\partial \underline{v}} = -A(\underline{\omega}) = \underset{\substack{\uparrow \\ \downarrow}}{\overset{\substack{\text{m} \\ \text{n}}}{\text{m}}} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \\ \frac{\partial \underline{r}(\underline{v}, \underline{\omega})}{\partial \underline{\omega}} = -DA(\underline{\omega}) \cdot \underline{v} = \underset{\substack{\uparrow \\ \downarrow}}{\overset{\substack{\text{p} \\ \text{m}}}{\text{m}}} \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \end{array} \right. \quad \begin{array}{l} \text{(V-17)} \\ \text{(V-18)} \end{array}$$

où $DA(\underline{\omega})$ est un tenseur à 3 dimensions $m \times n \times p$. Compte tenu de la définition de $I(\underline{v}, \underline{\omega})$, on en déduit immédiatement :

$$\text{gradient de } I \left\{ \begin{array}{l} \frac{\partial I(\underline{v}, \underline{\omega})}{\partial \underline{v}} = -2\underline{r}^T A(\underline{\omega}) = 1 \updownarrow \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \\ \frac{\partial I(\underline{v}, \underline{\omega})}{\partial \underline{\omega}} = -2\underline{r}^T \cdot DA(\underline{\omega}) \cdot \underline{v} = 1 \updownarrow \begin{array}{|c|} \hline \text{ } \\ \hline \end{array} \end{array} \right. \quad \begin{array}{l} \text{(V-19)} \\ \text{(V-20)} \end{array}$$

Considérons maintenant le critère projeté $\tilde{I}(\underline{\omega})$ où $\underline{v}$ est devenu une fonction de $\underline{\omega}$. Rappelons que :

$$\tilde{I}(\underline{\omega}) = \min_{\underline{v}} I(\underline{v}, \underline{\omega}) = \min_{\underline{v}} \|\underline{y} - A(\underline{\omega}) \underline{v}\|^2$$

Il est bien connu que la solution $\hat{\underline{v}}(\underline{\omega})$ du problème de moindres carrés est donnée par :

$$\hat{\underline{v}} = A^+(\underline{\omega}) \underline{y} \quad \text{(V-21)}$$

où A^+ est la pseudo inverse de A au sens de Penrose [19]. Par suite :

$$\begin{aligned} \hat{\underline{r}} &= \underline{y} - AA^+ \underline{y} \\ \hat{\underline{r}} &= (\mathbf{1} - AA^+) \underline{y} \end{aligned} \quad \text{(V-22)}$$

Définissons P_A comme étant l'opérateur projection orthogonale sur l'espace engendré par les colonnes de A , et $P_A^\perp$ l'opérateur projection orthogonale sur le complément orthogonal de l'espace précédent. On sait alors que :

$$\begin{aligned} P_A &= AA^+ \\ P_A^\perp &= \mathbf{1} - P_A \end{aligned} \quad \text{(V-23)}$$

relation associée à l'interprétation géométrique bien connue des moindres carrés (Fig. 6).

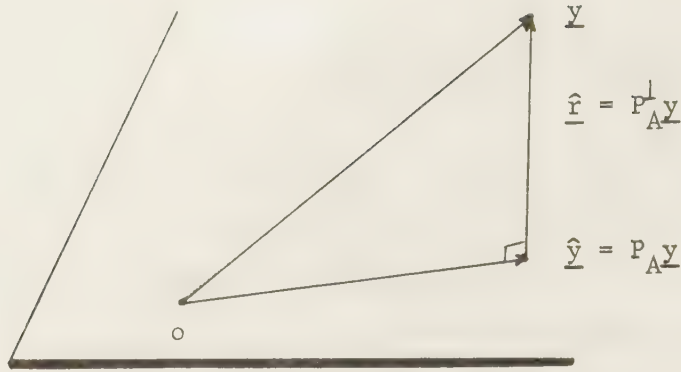


Figure 6

De (V-22) et (V-23), on en déduit finalement l'expression :

$$\tilde{I}(\underline{\omega}) = ||P_{A(\underline{\omega})}^{\perp} \cdot \underline{y}||^2 \quad (V-24)$$

... à partir de laquelle devront être calculés Jacobien et gradient. Au préalable, nous énoncerons les deux lemmes suivants :

Lemme V-1

Soit $A^+(\underline{\omega})$ une matrice $m \times n$ telle que $AA^+A = A$ et $(AA^+)^T = AA^+$,

alors :

$$DP_A = P_A^{\perp} \cdot DA \cdot A^+ + (P_A^{\perp} \cdot DA \cdot A^+)^T \quad (V-25)$$

Démonstration : voir [19]

Remarque : De (V-23) il vient $DP_A^{\perp} = -DP_A$

Lemme V-2

Soit A^+ la pseudo inverse de A au sens de Penrose. Alors :

$$P_A^{\perp} (A^+)^T = 0 \quad (V-26)$$

Démonstration :

$$\begin{aligned} P_A^{\perp} (A^+)^T &= (I - AA^+) (A^+)^T = (A^+)^T - (AA^+)^T (A^+)^T \\ &= (A^+ - A^+ AA^+)^T = (A^+ - A^+)^T \\ &= 0 \end{aligned}$$

On est maintenant en mesure de calculer le Jacobien de $\underline{\hat{r}}$.

Reprenons (V-22), il vient :

$$\begin{aligned}\frac{\partial \underline{\hat{r}}(\underline{\omega})}{\partial \underline{\omega}} &= D \underline{P}_A^{\perp}(\underline{\omega}) \cdot \underline{y} \\ \frac{\partial \underline{\hat{r}}(\underline{\omega})}{\partial \underline{\omega}} &= - D \underline{P}_A(\underline{\omega}) \cdot \underline{y} \\ \frac{\partial \underline{\hat{r}}(\underline{\omega})}{\partial \underline{\omega}} &= - \left[\underline{P}_A^{\perp} \cdot DA \cdot A^+ + (\underline{P}_A^{\perp} \cdot DA \cdot A^+)^T \right] \underline{y} \\ \frac{\partial \underline{\hat{r}}(\underline{\omega})}{\partial \underline{\omega}} &= - \left[\underline{P}_A^{\perp} \cdot DA \underline{\hat{v}} + (A^+)^T (DA)^T \underline{\hat{r}} \right]\end{aligned}\tag{V-27}$$

Cette expression relativement compliquée peut néanmoins être calculée très efficacement au moyen toujours des transformations de Householder. On renvoie le lecteur en [20] pour la description précise d'une mise en oeuvre numérique.

Le gradient de $\tilde{I}$ se déduit maintenant facilement de (V-27) :

$$\begin{aligned}\frac{\partial \tilde{I}(\underline{\omega})}{\partial \underline{\omega}} &= 2 \underline{y}^T \underline{P}_A^{\perp} \cdot \frac{\partial \underline{\hat{r}}}{\partial \underline{\omega}}(\underline{\omega}) \\ &= 2 \underline{y}^T \underline{P}_A^{\perp} DA \underline{\hat{v}} - \underline{y}^T \underline{P}_A (A^+)^T (DA)^T \underline{\hat{r}}\end{aligned}$$

qui d'après (V-26), se réduit à :

$$\frac{\partial \tilde{I}(\underline{\omega})}{\partial \underline{\omega}} = - 2 \underline{\hat{r}} DA \underline{\hat{v}}\tag{V-28}$$

En d'autres termes, on a le résultat, important en pratique :

$$\frac{\partial \tilde{I}(\underline{\omega})}{\partial \underline{\omega}} = \left. \frac{\partial I(\underline{v}, \underline{\omega})}{\partial \underline{\omega}} \right|_{\underline{v}=\underline{\hat{v}}}\tag{V-29}$$

2.2.3 - Application au critère erreur de sortie

On peut envisager 4 niveaux pour traiter le problème d'optimisation de l'erreur de sortie :

niveau_1 - minimiser directement $I(\underline{v}, \underline{\omega})$ - utilisation de ∇I (gradient)

niveau_2 - minimiser $I(\underline{v}, \underline{\omega})$ par régression non linéaire - utilisation de $\nabla_{\underline{\omega}}$ (Jacobien)

niveau 3 - minimiser directement $\hat{I}(\underline{\omega})$ - utilisation de $\nabla \hat{I}$ (gradient)

niveau 4 - minimiser $\hat{I}(\underline{\omega})$ par régression non linéaire - utilisation de $\nabla \hat{I}$ (Jacobien)

Dans le cas présent, pour calculer les quantités I , $\hat{I}$, ∇I , $\nabla \hat{I}$, $\nabla \hat{E}$, $\nabla \hat{E}$, il convient donc de préciser ce qu'on a désigné par $A(\underline{\omega})$ et $DA(\underline{\omega})$, éléments de base des équations (V-17) à (V-20), (V-27) et (V-28).

- Calcul de A

Pour cela revenons à la récurrence (V-11), que l'on récrit ici :

$$\eta_t = a_1 \eta_{t-1} + \dots + a_n \eta_{t-n} + b_0 u_t + \dots + b_n u_{t-n}$$

Notons $\underline{\eta}$ le vecteur de composante η_t , $t = n, \dots, N$, il s'agit alors de définir une matrice A telle que :

$$\underline{\eta} = A \underline{\alpha} \quad (V-30)$$

$$\text{avec } \underline{\alpha}^T = (\eta_{n-1}, \eta_{n-2}, \dots, \eta_0 \mid b_n, b_{n-1}, \dots, b_0)$$

De (V-11), il vient pour $t = n$

$$\eta_n = [a_1, a_2, \dots, a_n \mid u_0, u_1, \dots, u_n] \cdot \underline{\alpha} \triangleq \underline{\ell}_1^T \underline{\alpha}$$

pour $t = n + 1$

$$\begin{aligned} \eta_{n+1} &= [a_2, \dots, a_n, 0 \mid u_1, u_2, \dots, u_{n+1}] \underline{\alpha} + a_1 \eta_n \\ &= (\underline{\ell}_2^T + a_1 \underline{\ell}_1^T) \underline{\alpha} \triangleq (\underline{\ell}_2')^T \underline{\alpha} \end{aligned}$$

pour $t = n + 2$, il vient :

$$\begin{aligned} \eta_{n+2} &= [a_3, \dots, a_n, 0, 0 \mid u_2, \dots, u_{n+2}] \underline{\alpha} + a_1 \eta_{n+1} + a_2 \eta_n \\ &= \underline{\ell}_3^T \underline{\alpha} + a_1 \underline{\ell}_2'^T \underline{\alpha} + a_2 \underline{\ell}_1^T \underline{\alpha} \\ &\triangleq \underline{\ell}_3'^T \underline{\alpha} \end{aligned}$$

.

puis pour $t = 2n$:

$$\begin{aligned} \eta_{2n} &= [0, \dots, 0 \mid u_n, \dots, u_{2n}] \underline{\alpha} + a_1 \eta_{2n-1} + \dots + a_n \eta_n \\ &= \underline{\ell}_{n+1}^T \underline{\alpha} + a_1 \underline{\ell}_n'^T \underline{\alpha} + \dots + a_n \underline{\ell}_1^T \underline{\alpha} \\ &\triangleq \underline{\ell}_{n+1}'^T \underline{\alpha} \end{aligned}$$

et enfin pour $t \geq 2n$, avec $\tau = t - n + 1$, on obtient :

$$\begin{aligned} \eta_t &= [0, \dots, 0 \overset{n}{\mid} u_{\tau-1}, \dots, u_t] \underline{\alpha} + a_1 \eta_{t-1} + \dots + a_n \eta_{t-n} \\ &= \ell_{\tau}^T \underline{\alpha} + a_1 \ell_{\tau-1}^T \underline{\alpha} + \dots + a_n \ell_{\tau-n}^T \underline{\alpha} \\ &\triangleq \ell_{\tau}^T \underline{\alpha} \end{aligned}$$

En résumé, on a donc :

$$\eta_t = \ell_{\tau}^T \underline{\alpha} \quad t = n, \dots, N \quad (V-31)$$

avec :

$$\begin{cases} \ell_{\tau}^T = \ell_{\tau}^T + a_1 \ell_{\tau-1}^T + \dots + a_j \ell_{\tau-j}^T \\ \ell_1^T = \ell_1^T \end{cases} \quad \text{où } j = \min(\tau-1, n) \quad (V-32)$$

et

$$\ell_{\tau} = \begin{cases} [a_{\tau}, \dots, a_n, \overset{\tau-1}{0}, \dots, 0 \mid u_{\tau-1}, \dots, u_t] & \tau = 1, \dots, n \\ [0, \dots, 0 \mid u_{\tau-1}, \dots, u_t] & \tau > n \end{cases} \quad (V-33)$$

... de sorte que la matrice $A(\underline{a})$, on a noté $\underline{a}^T \Delta(a_1, \dots, a_n)$, se construit ligne par ligne au moyen d'une simple récurrence : (V-32).

- Calcul de DA

Il se déduit directement de (V-32) et (V-33) d'où on tire :

$$\frac{D\ell_{\tau}^T}{\partial a_i} = \frac{D\ell_{\tau}^T}{\partial a_i} + a_1 \frac{D\ell_{\tau-1}^T}{\partial a_i} + \dots + a_j \frac{D\ell_{\tau-j}^T}{\partial a_i} + \ell_{\tau-i}^T \quad (V-34)$$

$\tau = 1, \dots, N-n+1$

avec la convention $\ell_k^T = 0$ si $k \leq 0$

et par ailleurs :

$$\frac{D\ell_{\tau}^T}{\partial a_i} = \begin{cases} \overset{i-\tau}{[0 \dots 0 \ 1 \ 0 \ \dots \ 0]} & i - \tau > 0 \\ \xrightarrow{2n+1} & \\ [0 \ \dots \ \dots \ 0] & i - \tau \leq 0 \end{cases} \quad (V-35)$$

Mise en oeuvre de la procédure :

Sans entrer dans les détails, nous voudrions faire ici quelques remarques d'ordre pratique, concernant cette approche de l'optimisation de l'erreur de sortie.

Tout d'abord il convient de noter qu'il n'est pas nécessaire de stocker complètement en mémoire A et DA. En effet ces matrices (bi et tri-dimensionnelles) doivent être multipliées à droite par $\underline{U}$ ou/et à gauche par $\underline{\varepsilon}_n$, opérations qui peuvent être effectuées au fur et à mesure que les lignes de A et DA sont calculées. En raison de la nature récursive de ces calculs, équations (V-32) et (V-34), il suffit en fait de ne mémoriser que n lignes consécutives.

Du point de vue numérique, le calcul de A et par voie de conséquence de DA peut par ailleurs présenter quelques difficultés. De fait lorsque les coefficients a_i avec lesquels sont effectués ces calculs correspondent à des pôles, en module, supérieurs à 1 (système instable), les éléments des lignes $\underline{u}_t$ croissent rapidement et peuvent facilement atteindre le dépassement de capacité ("overflow"). Il est donc indispensable de se prémunir contre ce risque; on peut envisager deux possibilités pour y parvenir :

- interrompre le calcul dès que les éléments de A (et DA) atteignent, en valeur absolue, un certain seuil, 10^{20} par exemple, ou bien utiliser le contrôle d'"overflow".
- une deuxième approche consiste à tester la stabilité des pôles du système avant d'effectuer tout autre calcul. On signalera à cet effet une procédure particulière simple à mettre en oeuvre, due à Astrom Jury et Agniel [21] et basée sur le théorème de Schur-Cohn. Notons qu'à la limite, on pourrait également calculer les pôles du système, ce qui est également réalisable très efficacement [22].

Notons, par ailleurs, que le risque inverse d'"underflow" (pôle voisin de zéro en module) est beaucoup moins grave, puisque la plupart des compilateurs remplace le nombre en question par zéro. Il convient néanmoins d'inhiber le message d'erreur correspondant s'il y a lieu.

Indiquons pour terminer, qu'il est avantageux de traiter séparément les deux cas particuliers où $\{u_t\}$ est respectivement l'entrée impulsion et l'entrée en échelon. Dans ces hypothèses la construction de A et DA se trouve notablement simplifiée et il est encore possible de réduire l'encombrement mémoire nécessaire. Le lecteur trouvera en [23] un développement détaillé de la mise en oeuvre de la technique du critère projeté, appliquée à ce type d'entrées/sorties.

2.3 - Estimation de l'ordre d'un modèle

La détermination de l'ordre d'un modèle est un problème difficile en soi. Il n'existe pas, en effet, de critère universel, répondant aux exigences des différents contextes, et capable d'indiquer si la complexité d'un système est significative vis-à-vis des informations à partir desquelles il a été défini.

On rejoint dans ce domaine la notion de vérification ou validation de modèle qui, là encore, ne peut être résolu dans son ensemble. Au mieux peut-on juger de la qualité d'un modèle dans un but précis. Ceci explique les nombreuses approches théoriques basées, soit sur des considérations statistiques (test d'hypothèses), soit déterministes [24],[25],[26],[27]. Malgré leur contexte individuel, quelquefois assez restrictif, elles n'en représentent pas moins un support précieux sur la base duquel il convient de tirer des conclusions appropriées à chaque cas particulier.

L'expérience est là pour confirmer la quasi nécessité d'introduire "dans la boucle" une intervention humaine, réalisant une opération de synthèse, ici précisément difficile à systématiser [28],[29].

Notre propos n'en est pas pour autant de reprendre d'une manière exhaustive un ensemble possible d'approches, cf [25],[26], en particulier celles basées sur des hypothèses probabilistes. Nous avons plutôt préféré nous en tenir à quelques réflexions d'ordre théorique et pratique relatives principalement aux critères d'approximation précédemment introduits et dont la généralité d'application prévaut souvent sur des solutions plus sophistiquées.

2.3.1 - Erreur d'équation et ordre du système

Rappelons qu'un système d'ordre n , identifiable à partir des entrées/sorties $\{u_t, y_t\}$ $t = 0, \dots, N$ est caractérisé par les relations (V-5) et (V-6) qui sont (pour mémoire) :

$$\begin{cases} \underline{y}_n = \tilde{R}_{n,N} \underline{\alpha}_n & \text{(V-5)} \\ \underline{y}_k \neq \tilde{R}_{k,N} \underline{\alpha}_k & k < n \end{cases} \quad \text{(V-6)}$$

Par définition le modèle approché au sens du critère erreur d'équation est obtenu en résolvant (V-6) au sens des moindres carrés. Soit $\underline{\alpha}_k^*$ la solution correspondante et $||\underline{e}_k^*||$ la norme de l'erreur d'équation associée. Il est alors clair que $||\underline{e}_k^*||$ est une fonction non croissante de k qui de surcroît est nulle pour tout $k > n$. En d'autres termes l'ordre du système se caractérise comme étant la première valeur de k pour laquelle on a $||\underline{e}_k^*|| = 0$. Pour des raisons évidentes, celle-ci ne pourra jamais s'annuler numériquement, mais en fait sera considérée comme telle lorsque :

$$||\underline{e}_k^*|| < \eta ||\underline{y}_k|| \quad \text{(V-36)}$$

où η est la précision (relative) de la machine.

Lorsqu'il s'agit de réduire l'ordre d'un système ou de développer un modèle approché d'un processus physique, il convient de substituer à η

une quantité ζ positive, plus ou moins petite, représentative de l'erreur d'équation tolérée. La difficulté pratique réside alors dans le choix d'une valeur appropriée de ζ , $||\underline{e}_k^*||$ n'ayant pas la signification physique directe d'une erreur de sortie. En fait, la détermination de n devra se faire non seulement sur la base de l'inégalité :

$$||\underline{e}_n^*|| < \zeta ||\underline{y}_n|| \quad (V-37)$$

mais encore sur la nature de l'évolution de $||\underline{e}_k^*||$ en fonction de k , et éventuellement sur l'introduction de pôles et zéros voisins. La Fig. 7, par exemple, illustre le cas du choix d'une valeur de ζ correspondant à un critère trop sévère par rapport aux données du problème (ordre n') alors qu'intuitivement on peut penser que la complexité d'un système d'ordre supérieur à n n'est que peu significative ici.

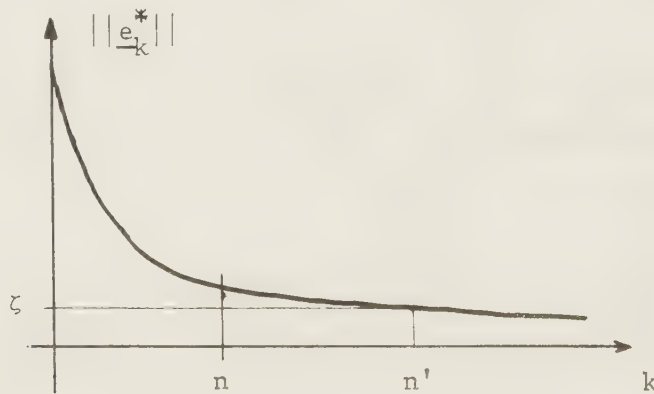
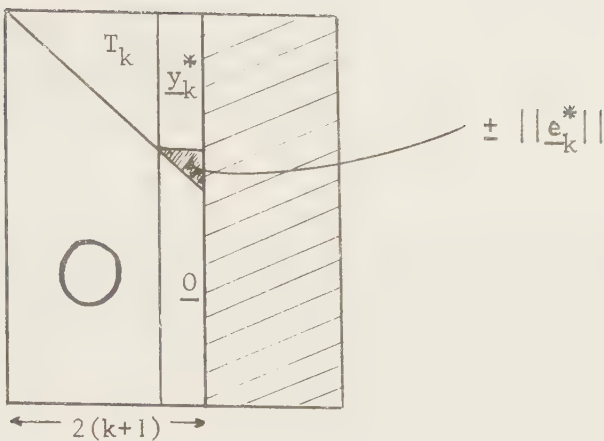


Figure 7

L'intérêt de cette approche réside essentiellement dans la simplicité avec laquelle on peut calculer $||\underline{e}_k^*||$ en fonction de k , et ceci *sans* avoir à déterminer pour autant les paramètres correspondants du modèle $\underline{\alpha}_k^*$. En effet, supposons les données $\{u_t, y_t\}$ disposées suivant la matrice R_{qN} (Cf. V-3), où q est une borne supérieure pour n , connue. Soit donc :

$$R_{qN} = \left[\begin{array}{cc|cccc|cc} u_0 & y_0 & & & & & u_q & y_q \\ \vdots & \vdots & & & & & \vdots & \vdots \\ \cdot & \cdot & & & & & \cdot & \cdot \\ \cdot & \cdot & & & & & \cdot & \cdot \\ \cdot & \cdot & & & & & \cdot & \cdot \\ \hline u_{N-q} & y_{N-q} & & & & & u_N & y_N \end{array} \right] \quad (V-38)$$

et appliquons aux colonnes de R_{qN} une suite de transformations élémentaires de Householder (Cf Ch IV § 2) telles que :

$$Q_k R_{qN} =$$


(V-39)

On a alors :

$$\underline{\alpha}_k^* = T_k^{-1} \underline{y}_k^* \quad (V-40)$$

La démonstration en est très simple. Considérons la résolution du système $\underline{Ax} = \underline{b}$ au sens des moindres carrés où A est une matrice $m \times n$ ($m > n$) de rang n . Soit Q une matrice orthogonale $m \times m$, il vient alors :

$$||\underline{Ax} - \underline{b}|| = ||Q\underline{Ax} - Q\underline{b}|| \triangleq ||\tilde{\underline{Ax}} - \tilde{\underline{b}}||$$

Par ailleurs on sait qu'il existe une matrice Q telle que [10] :

$$\tilde{A} = \begin{bmatrix} T \\ 0 \end{bmatrix} \quad \text{et} \quad \tilde{b} = \begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \end{bmatrix}$$

Il en découle :

$$||\tilde{\underline{Ax}} - \tilde{\underline{b}}|| = ||T\underline{x} - \underline{b}_1|| + ||\underline{b}_2||$$

Il est clair que le minimum de $||\underline{Ax} - \underline{b}||$ est atteint pour $\underline{x} = T^{-1} \underline{b}_1$ et vaut $||\underline{b}_2||$. La transformation des notations au problème posé démontre donc (V-40) et que l'erreur d'équation $||e_k^*||$ est au signe près égale aux terme diagonaux des colonnes de sorties $\underline{y}_k$ après transformation (V-39).

En d'autres termes, au cours de la triangularisation de R_{qN} apparaît sur les termes diagonaux de rang $2(k+1)$ la valeur du critère (au signe près) correspondant à un modèle d'ordre k sans avoir dû calculer les coefficients correspondants $\underline{\alpha}_k^*$, dont la solution exprimée en (V-40) est de toute évidence indépendante de la connaissance de $||e_k^*||$.

De ce fait, il est donc peu coûteux de cerner un intervalle n_0, n_1 pour l'ordre du système, en opérant ainsi, résultat que l'on pourra ensuite préciser en observant, par exemple, le comportement correspondant de l'erreur de sortie.

2.3.2 - Relation entre erreur d'équation et erreur de sortie

Celle-ci est immédiate à déterminer si on se réfère aux figures 3 et 4 où sont respectivement définies erreur d'équation et erreur de sortie. Reprenons la Fig. 7, il vient Fig. 8 :

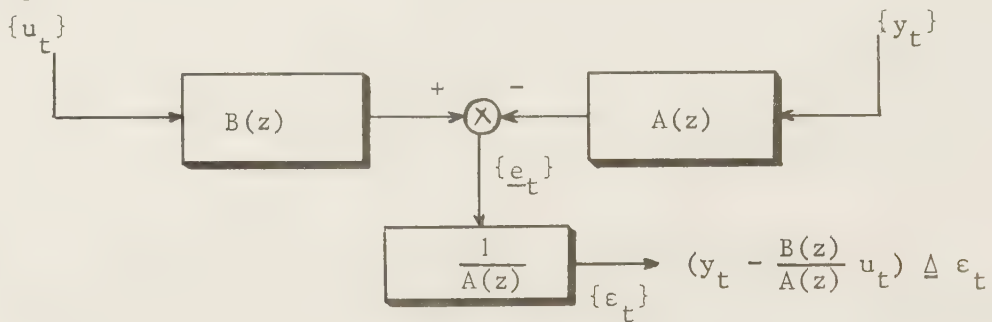


Figure 8

où $B(z) = b_0 + b_1 z^{-1} + \dots + b_n z^{-n}$
 $A(z) = 1 - a_1 z^{-1} - \dots - a_n z^{-n}$

Par suite de la Fig. 8 découle la relation cherchée :

$$\epsilon_t = a_1 \epsilon_{t-1} + \dots + a_n \epsilon_{t-n} + e_t \quad (V-41)$$

qui ne fait qu'explicitier le schéma fonctionnel de la Fig. 9.

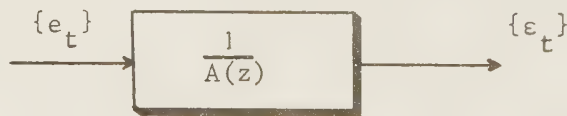


Figure 9

Notons $\hat{\alpha}_k$ la solution minimisant $||\epsilon_k(\alpha_k)|| \Delta ||\hat{\epsilon}_k||$, par suite on a :

$$||\epsilon_k(\alpha_k^*)|| \geq ||\epsilon_k(\hat{\alpha}_k)|| \quad (V-42)$$

Par ailleurs notons que :

$$||e_k^*(\alpha_k^*)|| \geq ||e_{k+1}^*(\alpha_{k+1}^*)||$$

n'implique pas nécessairement :

$$||\epsilon_k(\alpha_k^*)|| \geq ||\epsilon_{k+1}(\alpha_{k+1}^*)|| \quad (V-43)$$

On n'est seulement assuré que l'inégalité (V-44) est vérifiée.

$$||\underline{\varepsilon}_k(\hat{\alpha}_k)|| \geq ||\underline{\varepsilon}_{k+1}(\hat{\alpha}_{k+1})|| \quad (V-44)$$

On est donc amené à réaliser un compromis entre l'utilisation d'une information précise (physiquement) telle que l'erreur de sortie $||\hat{\underline{\varepsilon}}_k||$ mais relativement lourde à obtenir, et la connaissance de l'erreur d'équation $||\underline{e}_k^*||$ rapidement calculée mais d'une interprétation moins immédiate.

Dans ce sens, l'utilisation mixte de ces deux critères, telle qu'on l'a suggéré un peu plus haut est un moyen de répondre à cet objectif[†].

La détermination de l'ordre d'un système a reposé jusqu'ici sur des critères relevant de propriétés externes (entrées/sorties). Nous allons maintenant discuter d'un autre aspect du problème sous l'angle de la réductibilité du modèle d'état. Plus précisément, nous nous proposons de voir quelles informations supplémentaires peuvent être obtenues en raisonnant sur le degré de réductibilité d'un modèle en termes d'observabilité ou/et d'accessibilité.

2.3.3 - Pseudo réductibilité

Jusqu'alors les notions d'accessibilité, d'observabilité et par suite de réductibilité ont été définies comme des problèmes "tout ou rien". En effet, par analogie avec $||\underline{e}_k^*||$ ou/et $||\hat{\underline{\varepsilon}}_k||$ qui ne s'annulent théoriquement, pour la première fois, que pour $k = n$, (l'ordre du système), les matrices d'accessibilité ou/et d'observabilité ne sont de rang maximum k que pour $k < n$. Néanmoins, en pratique, traitement sur calculateur numérique, réduction de système, approximation de processus réel, ces conditions de rang ne peuvent s'exprimer d'une manière aussi tranchée. On est ramené une fois de plus à déterminer quand un vecteur est linéairement dépendant, ou non, d'un ensemble d'autres vecteurs.

Nous ne considérerons dans la suite que l'optique accessibilité, en raison d'une part, de la **dualité** qui lie ce concept à l'observabilité, et d'autre part, **parce** que dans le cas général de la réalisation d'un système à partir d'entrées/sorties quelconques, on a vu que le modèle canonique utilisé était nécessairement observable.

† Une alternative étant d'opérer avec $||\hat{\underline{\varepsilon}}_k||$ dans les mêmes conditions que $||\underline{e}_k^*||$

Il s'agit donc, en fait de substituer au concept classique d'accessibilité une notion exprimant plutôt un degré d'accessibilité, mesuré par un critère approprié, par exemple l'énergie dissipée pour atteindre un état quelconque, noté $\underline{x}_N$.

Nous allons montrer que l'énergie minimum, correspondant au cas le plus défavorable (état le plus difficile à atteindre) est directement lié au conditionnement de la matrice d'accessibilité ; résultat dont on peut déduire une procédure de réduction basée sur le degré de mauvais conditionnement de cette matrice. Cette solution est en fait équivalente à la technique d' ϵ - réductibilité introduite par FURUTA [34] dans le cas où $N = n$.

On peut donc voir dans les développements qui suivent une interprétation physique de cette "pseudo réductibilité" [34] considérée dans un cadre plus général : $N \geq n - \rho(\Gamma)$.

Revenons au critère d'énergie, mentionné plus haut, et exprimé de la manière suivante :

$$I \triangleq \sum_{i=0}^{N-1} \underline{u}_i^T \underline{u}_i = \underline{u}^T \underline{u} = ||\underline{u}||^2 \quad (V-45)$$

en posant :

$$\underline{u}^T = [\underline{u}_0^T, \underline{u}_1^T, \dots, \underline{u}_{N-1}^T] \quad (V-46)$$

Explicitons maintenant $\underline{x}_N$, l'accessibilité étant définie pour un état initial nul; il vient :

$$\underline{x}_N = [\Phi^{N-1}\Gamma, \dots, \Phi\Gamma, \Gamma]\underline{u} \quad (V-47)$$

$$\underline{\triangle} F \underline{u}$$

Admettons que le système (Φ, Γ) soit accessible, au sens classique du terme, c'est-à-dire $\rho(F) = n$, alors la commande $\underline{u}$ (V-46) minimisant I (V-45) est donnée par :

$$\underline{\hat{u}} = F^+ \underline{x}_N \quad (V-48)$$

où F^+ est la pseudo inverse de F au sens de Penrose [19]. Par suite on aura :

$$\hat{I} = ||\underline{\hat{u}}||^2 = ||F^+ \underline{x}_N||^2 = ||\underline{x}_N|| (F^{+T} F^+) \quad (V-49)$$

Avec l'hypothèse $\rho(F) = n$, on a bien sûr :

$$F^+ = F^T (FF^T)^{-1}$$

et

$$F^{+T} F^+ = (FF^T)^{-1} \quad (V-50)$$

Désignons par $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ les valeurs singulières de la matrice $F(n \times Nm)$ qui sont également les racines carrées des valeurs propres de FF^T . Rappelons en deux propriétés :

$$\rho(F) = r \quad \sigma_r > 0 \text{ et } \sigma_{r+1} = \dots = \sigma_n = 0 \quad (V-51)$$

$$\sigma_i \text{ de } F^+ = \begin{cases} 1/\sigma_i & \text{si } \sigma_i \neq 0 \\ 0 & \text{si } \sigma_i = 0 \end{cases} \quad (V-52)$$

Introduisons enfin la norme d'une matrice (A par exemple) associée à la norme d'un vecteur :

$$||A|| = \max_{||\underline{x}||=1} ||A\underline{x}|| \quad (V-53)$$

Lorsque $||\underline{x}||$ désigne la norme euclidienne, ce que nous avons implicitement fait jusqu'ici, on a alors :

$||A|| = ||A||_S =$ plus grande valeur singulière de A
qui est ce qu'on appelle usuellement la norme *spectrale* de la matrice [31]

Appliquons maintenant ces résultats à notre problème. En supposant $||\underline{x}_N|| = 1$, condition à laquelle on peut se ramener, on en déduit que :

$$\max_{||\underline{x}_N||} \hat{I} = \frac{1}{\sigma_n^2} = \hat{I} \quad (V-54)$$

En d'autres termes au fur et à mesure que les lignes de F deviennent de moins en moins linéairement indépendantes, le système (V-47) devient de plus en plus mal conditionné, σ_n décroît et par voie de conséquence (V-54) le maximum de $\hat{I}$ croît. C'est ce phénomène qui se produit lorsqu'on augmente l'ordre k d'un modèle, la plus petite valeur singulière de F s'annulant théoriquement lorsque k passe de l'ordre exact n à n + 1. Pour k = n+2, les deux plus petites valeurs singulières seraient nulles etc... On conçoit donc facilement que l'évolution des valeurs singulières de F représente une information concernant l'ordre du système. Notons que cette matrice n'est autre que la matrice d'accessibilité du système (en toute rigueur pour N = n) ; par suite on énoncera la ...

Définition V-3 - Un système dynamique d'ordre n est dit ϵ - accessible sur un horizon $N > n$, si $\hat{I} < \frac{1}{\epsilon^2}$, ou ce qui est équivalent si $\sigma_n[\Phi^{N-1}\Gamma, \dots, \Gamma] > \epsilon$

On dira alors que :

Définition V-4 - Un système dynamique d'ordre n , dans une forme canonique observable est ε - irréductible, si il est ε - accessible.

En conséquence, lorsque l'on a $0 < \sigma_n < \varepsilon$ on dira que le système est *pseudo réductible*, plutôt que réductible, en raison du seuil introduit par ε .

Par ailleurs, il est clair qu'en posant $\varepsilon = 0$, on retrouve les concepts classiques d'accessibilité ($\sigma_n(F) > 0 \iff \rho(F) = n$) et d'irréductibilité.

L' ε - accessibilité définit donc, en quelque sorte, un autre critère pour réduire un modèle. Nous n'en avons pas parlé plus tôt (§ II-1) parce que d'une part, cette approche ne permet pas généralement de résoudre directement le problème d'approximation (car elle présuppose la connaissance de F), et d'autre part, les réalisations construites suivant ce critère ne seront pas nécessairement obtenues (directement) dans une forme canonique. Néanmoins nous proposons quelques aspects caractéristiques de cette voie, mettant en lumière le concept d' ε - accessibilité et les liens qu'on peut lui attribuer avec les critères précédents.

Proposition V-1

Etant donné un système dynamique d'ordre n , le modèle réduit d'ordre k , minimisant le maximum de $\hat{I}(\hat{I})$ est obtenu en annulant les $n-k$ plus petites valeurs singulières de F .

Démonstration - Elle découle de la relation (V-54)

Avant de poursuivre, introduisons les notations suivantes. Soit $U.\Omega.V^T$ la "décomposition en valeurs singulières" de F . Rappelons que par définition U et V sont des matrices composées des vecteurs propres orthonormés de FF^T et F^TF , et Ω est une matrice nulle à l'exception des termes diagonaux égaux à $\sigma_1, \dots, \sigma_n$.

On est maintenant en mesure d'énoncer la :

Proposition V-2

Etant donné un système dynamique $\Sigma_n(\Phi, \Gamma, H, D)$, désignons par U_k la matrice constituée des k premières colonnes de U associées aux k premières valeurs singulières de $F(\sigma_1, \dots, \sigma_k)$, alors le modèle réduit minimisant $\hat{I}$ est donné par :

$$\Phi_k = U_k^T \Phi U_k ; \quad \Gamma_k = U_k^T \Gamma ; \quad H_k = H U_k ; \quad D_k = D \quad (V-55)$$

Démonstration - D'après la proposition (V-1) le modèle réduit est obtenu en annulant les $n-k$ dernières valeurs singulières de F . Après cette opération on a alors : $\rho[\Phi^{n-1}\Gamma, \dots, \Gamma] = k < n$. D'après le théorème de réduction II-3, il existe une matrice A telle que :

$$A^{-1} \Phi A = \overset{k}{\updownarrow} \begin{bmatrix} \Phi_k' & // \\ 0 & // \end{bmatrix}; \quad A^{-1} \Gamma = \begin{bmatrix} \Gamma_k \\ 0 \end{bmatrix} \overset{k}{\updownarrow}; \quad HA = [H_k \quad //]$$

La matrice U peut précisément jouer ce rôle puisque ses k premières colonnes engendrent l'espace accessible. Du théorème II-3 découle alors (V-55).

Un cas particulier intéressant est celui où les données du problème se présentent sous forme d'une séquence de pondération $\{f_t\}$. En effet la matrice F exprimée dans la base correspondante à la forme canonique (2) (Cf. Ch. III) est alors directement connue puisqu'égal à :

$$F = \begin{bmatrix} f_1 & . & . & . & f_N \\ . & & & & . \\ . & & & & . \\ f_n & . & . & . & f_{N+n-1} \end{bmatrix} \quad (V-56)$$

Par suite, étant donné une séquence $\{f_t\}$, $t = 0, \dots, N$ et une borne supérieure de l'ordre du système, soit q , construisons la matrice F_{qN} définie par :

$$F_{qN} = \begin{bmatrix} f_1 & . & . & f_q \\ . & & & . \\ . & & & . \\ f_{N-q} & . & . & f_{N-1} \end{bmatrix} \triangleq [\underline{f}_1 \quad . \quad . \quad \underline{f}_q] \quad (V-57)$$

Remarquons alors que la solution $\underline{a}_k^*$ du système :

$$F_{kN} \underline{a}_k = \underline{f}_{k+1} \quad (V-58)$$

au sens des moindres carrés correspond à un modèle du type "erreur d'équation" dans l'hypothèse où $\{u_t\}$ est l'entrée impulsion.

On dispose donc ici de deux analyses explicites pour déterminer une approximation via un pseudo rang de matrice. La première consiste à triangulariser F_{qN} au moyen des transformations de Householder :

$$Q^F_{qN} = \begin{array}{c} \begin{array}{|c|} \hline \begin{array}{c} \text{Diagram of a square matrix of size } k+1 \times k+1. \text{ The matrix is partitioned into a } k \times k \text{ block (top-left), a } k \times 1 \text{ block (top-right), a } 1 \times k \text{ block (bottom-left), and a } 1 \times 1 \text{ block (bottom-right). The } k \times k \text{ block contains several circles. The } 1 \times 1 \text{ block contains a circle. The } k \times 1 \text{ block contains a circle. The } 1 \times k \text{ block contains a circle. The } 1 \times 1 \text{ block contains a circle. The } k \times k \text{ block is labeled } S_k. \end{array} \\ \hline \end{array} \end{array} \quad (V-59)$$

Les paramètres du modèle d'ordre k sont alors solution de :

$$T_{k-k}^* = f_{k+1}^* \quad (V-60)$$

auquel correspond le critère $||e_k^*||$. On raisonne donc comme si S_k était négligeable ou nul, ce qui revient à considérer que F_{qN} était de rang k .

La seconde voie consiste à calculer la décomposition en valeurs singulières de F_{qN} , soit donc :

$$F_{qN} = V\Omega U^T \text{ avec } \Omega = \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_k & & 0 \\ & & & \ddots & \\ 0 & & & & \sigma_q \\ & & & & & 0 \end{bmatrix} \quad (V-61)$$

Cette opération s'effectue au moyen de transformations de Householder et par application de l'algorithme du Q.R. dont la stabilité est bien connue [33].

Dans ces conditions la matrice F_{qN} est considérée de rang k lorsque $\sigma_{k+1}, \dots, \sigma_q$ sont négligeables au même titre que pourraient l'être $||e_{k+1}^*||, \dots, ||e_q^*||$. Plus précisément, notons :

$$\tilde{F}_k = V \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_k & & 0 \\ & & & \ddots & \\ 0 & & & & 0 \end{bmatrix} U^T \triangleq V \Omega_k U^T \quad (V-62)$$

la matrice obtenue en annulant les valeurs singulières considérées comme négligeable. Notons que cette matrice est la meilleure approximation (au sens de la norme euclidienne) de F_{qN} par une matrice de rang k . De plus on a : $||\tilde{F}_k - F_{qN}|| = (\sigma_{k+1}^2 + \dots + \sigma_q^2)^{1/2}$, [32].

Effectuons maintenant le changement de base U , cf (V-55) de sorte que le système à résoudre initialement au sens des moindres carrés (V-58) devient :

$$V \cdot \Omega_k \underline{a}' = \underline{f}_{q+1} \quad (V-63)$$

$$\Rightarrow \Omega_k \underline{a}' = \underline{g} \quad (V-64)$$

avec : $\underline{g} = V^T \underline{f}_{q+1}$ et $\underline{a} = U \underline{a}'$

La solution de (V-64) est alors immédiate :

$$\underline{a}' = \begin{bmatrix} g_1/\sigma_1 \\ \vdots \\ g_k/\sigma_k \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \xrightarrow{q} \underline{\Delta} \begin{bmatrix} \underline{a}_k \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (V-65)$$

d'où découlent les paramètres du système réduit (V-55) exprimé dans la forme canonique (2), soit :

$$\Phi_k = \begin{bmatrix} 0 & 1 & & \\ & & \ddots & 1 \\ & & & \\ -\underline{a}_k & \dots & -\underline{a}_1 \end{bmatrix} ; \quad \Gamma_k = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix} \quad H = [1 \ 0 \ \dots \ 0] \quad (V-66)$$

Nous résumons maintenant les aspects caractéristiques de ces deux analyses en notant que :

$$\begin{aligned} - \min_{\underline{a}_k} ||F_{kN-k} \underline{a}_k - \underline{f}_{k+1}|| &= ||F_{kN-k}^* \underline{a}_k - \underline{f}_{k+1}|| = ||\underline{e}_k^*|| \\ - \begin{cases} \min_{\underline{a}_q} ||\tilde{F}_{k-q} \underline{a}_q - \underline{f}_{q+1}|| &= ||\tilde{F}_k \underline{a}_q - \underline{f}_{q+1}|| = \left(\sum_{j=k+1}^{N-q} g_j^2 \right)^{1/2} \\ ||\underline{a}_q|| &\text{minimum} \end{cases} \end{aligned}$$

après avoir posé $\underline{\tilde{a}}_q = U \underline{a}'$, résultats que l'on interprète ainsi : l'approche "erreur d'équation" ($\equiv$ moindres carrés classiques $\equiv$ régression linéaire) consiste à résoudre un système surabondant de rang maximum (V-58) alors que le critère d' ϵ - accessibilité (ou énergie minimum $\hat{J}$) revient à trouver la solution de norme minimale d'un système surabondant contraint à être de rang (non maximum) égal à l'ordre du modèle approché (V-64).

Pour terminer cette analyse on notera la propriété suivante [43] : ...

2.3 - 2.4

Soit A une matrice $m \times n$ de rang n et R la matrice triangulaire obtenue par factorisation de A suivant les transformations de Householder, les éléments diagonaux étant rangés en module décroissant. Alors les valeurs singulières de A : $\sigma_1 > \sigma_2 > \dots, \sigma_n > 0$ et les éléments diagonaux de R (r_i) vérifient les inégalités :

$$2^{1-i} |r_i| < 3(4^i + 6i - 1)^{-1/2} |r_i| < \sigma_i < (n-i+1)^{1/2} |r_i| \quad (V-67)$$

... qui permet, sans avoir à effectuer la décomposition en valeurs singulières de F_{qN} , de connaître par simple triangularisation de F_{qN} , un intervalle contenant le maximum de $\hat{I}$.

2.4 - Extension à divers autres modèles

Dans l'hypothèse d'un système monovariable, on a montré que l'ordre d'un système pouvait être choisi sur la base d'un ensemble d'informations telles que l'erreur d'équation $||e_k^*||$, l'erreur de sortie $||\hat{e}_k||$ ou σ_k la $k^{\text{ième}}$ valeur singulière associée à une certaine matrice F dont est extrait la matrice d'accessibilité.

On a également vu, qu'en tenant compte de la structure particulière du critère $||e_k||$, son optimisation pouvait être réalisée par programmation non linéaire dans des espaces à k dimensions.

2.4.1 - Système multi-entrées

La transposition de l'ensemble de ces propriétés à des systèmes ayant plusieurs entrées ne présente aucune difficulté et se résume essentiellement à des changements de notations que nous allons simplement mentionner sans entrer dans les détails.

Considérons donc un système Σ_k d'ordre k , excité par m entrées indépendantes. Les entrées/sorties $\{\underline{u}_t, y_t\}$ sont alors reliées par la récurrence :

$$y_t = a_1 y_{t-1} + \dots + a_k y_{t-k} + \underline{b}_0^T \underline{u}_t + \dots + \underline{b}_{k-t}^T \underline{u}_{t-k} \quad (V-68)$$

que l'on peut encore récrire sous forme matricielle pour $t = k, \dots, N$:

$$\underline{y}_k = \tilde{R}_{k, N-k}^{\alpha} \quad (V-69)$$

seulement 8 interviennent non linéairement.

Enfin, en ce qui concerne les développements relatifs à l' ϵ -accessibilité, il suffirait de changer $\{f_t\}$ en $\{\underline{f}_t^T\} \triangleq [f_t^1, \dots, f_t^m]$

2.4.2 - Le modèle d'Astrom [35]

Ce modèle bien connu des automaticiens, se définit par la structure suivante Fig. 9 :

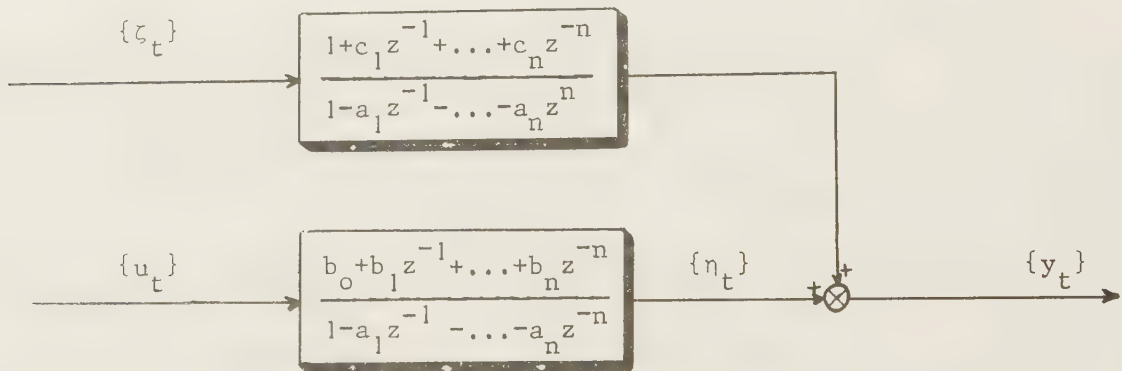


Figure 9

à laquelle correspond la récurrence :

$$y_t = a_1 y_{t-1} + \dots + a_n y_{t-n} + b_0 u_t + \dots + b_n u_{t-n} + \zeta_t + c_1 \zeta_{t-1} + \dots + c_n \zeta_{t-n} \quad (V-74)$$

sans vouloir introduire les justifications théoriques d'une telle structure, (perturbations gaussienne, maximum de vraisemblance), indiquons simplement que le problème posé consiste à minimiser l'écart quadratique moyen d'une suite $\{\xi_t\}$ défini par :

$$\xi_t = y_t - [a_1 y_{t-1} + \dots + a_n y_{t-n} + b_0 u_t + \dots + b_n u_{t-n} + c_1 \xi_{t-1} + \dots + c_n \xi_{t-n}]$$

$$t = n, \dots, N \quad (V-75)$$

où $\{u_t, y_t\}$ est ici la séquence d'entrées/sorties relative au processus objet.

Cette optimisation devra être effectuée par rapport aux $3n+1$ paramètres a_i, b_i, c_i auxquels on adjoint généralement les n conditions initiales $(\xi_0, \dots, \xi_{n-1})$. Notons qu'à l'optimum si modèle et processus sont identiques, les suites $\{\zeta_t\}$ et $\{\xi_t\}$ auront les mêmes caractéristiques (statistiques), Cf. (V-74) et (V-75).

Revenons à l'optimisation proprement dite de $||\underline{\xi}||$ où $\underline{\xi}^T = (\xi_n, \dots, \xi_N)$. Celle-ci, nous l'avons dit, doit être réalisée par rapport à $4n+1$ paramètres. Cependant, en observant l'expression (V-75) on constate facilement que ξ_t n'est une fonction non linéaire que des n paramètres c_i . Par suite en appliquant la technique du critère projeté, à ce problème, sa dimension pourra être réduite de $4n+1$ à seulement n , ce qui représente une alternative intéressante à expérimenter et à comparer avec la solution globale adoptée initialement par Astrom [35] en utilisant un algorithme du type Newton et d'ailleurs reprise plus tard par Gustavson [28],[30]).

Il est évident que ceci se généralise sans difficulté aux cas multi-entrées, avec des conclusions analogues au § 2.4.1.

Remarquons, enfin, que si l'on pose $c_i=0$ $i=1, \dots, n$, la méthode du modèle d'Astrom se réduit à l'approche par erreur d'équation, c'est-à-dire $||\underline{\xi}|| = ||\underline{e}||$, et dans le cas où l'on a $c_i=a_i$, $i=1, \dots, n$, le critère $||\underline{\xi}||$ est identique à celui de l'erreur de sortie $||\underline{\varepsilon}||$.

Parallèlement, rappelons que la méthode des moindres carrés (erreur d'équation) conduit à des résultats "biaisés"[†] en présence de perturbations indépendantes, sur la sortie des processus alors que cette approche devient optimale lorsque ces mêmes perturbations ont la corrélation d'un signal indépendant filtre par $1/(1-az^{-1}+\dots-a_n z^{-n})$.

Ces réflexions devraient expliquer pourquoi en pratique l'approche erreur d'équation appliquée à des systèmes réels dont les perturbations sont souvent corréllées, donne des résultats relativement bons en première approximation ; en particulier l'erreur de sortie correspondant à ces modèles $||\underline{\varepsilon}(\underline{\alpha}^*)||$ peut ne pas être très éloignée de l'optimum $||\underline{\hat{\varepsilon}}||$.

3 - RESULTATS EXPERIMENTAUX -

Rappelons que le lecteur ne devra pas s'attendre à trouver ici des développements expérimentaux sur bien des points qui, pourtant, présenteraient un intérêt certain. En effet une analyse qualitative approfondie des multiples aspects que présentent le problème d'approximation, tant du point de vue des méthodes numériques d'optimisation que de l'efficacité des différents critères introduits, dépasserait déjà très largement le cadre de ce travail, sans parler du cas des systèmes multivariabiles que l'on abordera § IV et V.

† L'estimation d'un paramètre est biaisée lorsque cette estimation présente une erreur systématique par rapport à la vraie valeur.

Nous avons donc fait un choix autour de ces deux thèmes essentiels en s'attachant à ne considérer que quelques conditions d'expérimentation types.

- Le premier thème (§ 3.1) concerne donc la comparaison de différentes techniques d'optimisation non linéaire, appliquées à la minimisation du critère d'erreur de sortie, dans le cas particulier de l'entrée impulsion (ajustement de la séquence de pondération). Ici seules seront considérées les performances des différents algorithmes mis en jeu, sans chercher à exploiter les résultats obtenus d'un point de vue identification. Ceci nous a permis de nous limiter à deux problèmes tests désignés sous le sigle "Pb₁" et "Pb₂".

Les données de "Pb1" se présentent sous la forme d'une transmittance en "z" du quatrième ordre :

$$Pb_1(z) = \frac{0. + 0.6z^{-1} - 0.4z^{-2} + 0.015z^{-3} - 0.109z^{-4}}{1 - 3z^{-1} + 3.33z^{-2} - 1.6z^{-3} + 0.2736z^{-4}}$$

dont on a calculé les 50 premiers points de sa séquence de pondération.

"Pb₂", quant à lui, est directement défini par une séquence de pondération, ou plus exactement par les 50 premiers points de cette séquence, dont l'allure correspond au type de réponse impulsionnelle observée expérimentalement sur un surchauffeur (réponse : ouverture soupape/température). On a reproduit au tableau 1 les valeurs numériques utilisées, la silhouette du graphe correspondant étant tracée Figure 10 en comparaison avec la réponse de "Pb₁".

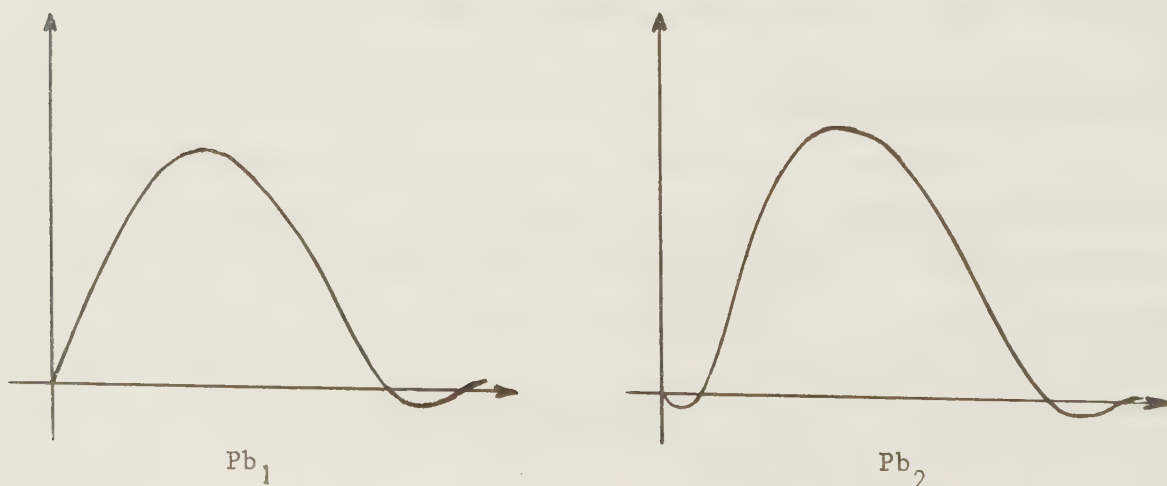


Figure 10

t	0	1	2	3	4	5	6	7	8	9
f _t	0	-0.9	-1	-0.75	-0.001	1.3	2.75	4.2	6.4	7.2
t	10	11	12	13	14	15	16	17	18	19
f _t	7.8	7.95	7.97	7.76	7.48	7.08	6.71	6.25	5.77	5.24
t	20	21	22	23	24	25	26	27	28	29
f _t	4.64	4.05	3.33	2.62	1.92	1.17	0.45	-0.02	-0.26	-0.35
t	30	31	32	33	34	35	36	37	38	39
f _t	-0.31	-0.17	0.	0.22	0.48	68	0.87	1.02	1.1	1.1
t	40	41	42	43	44	45	46	47	48	49
f _t	0.99	0.85	0.7	0.56	0.39	0.21	0.05	-0.08	-0.13	-0.16

Tableau 1

- Le deuxième thème (§ 3.2) se rapporte à l'identification d'un système monovariante du second ordre simulé suivant le schéma de la Fig. 11.

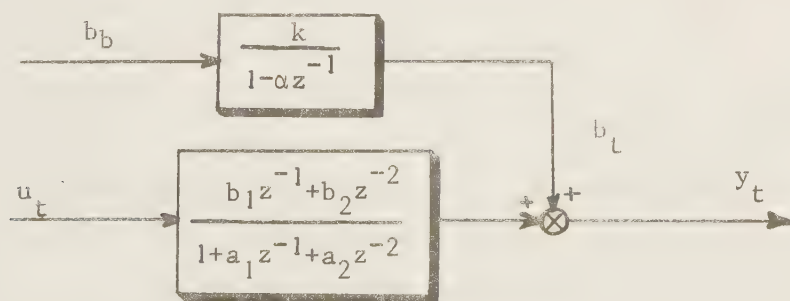


Figure 11

La perturbation b_t a été obtenue à l'aide d'une séquence indépendante gaussienne filtrée (paramètre α), et on a adopté comme signal d'entrée $\{u_t\}$ une autre séquence indépendante ; enfin le système, utilisé ici, est défini par les paramètres :

$$\begin{aligned} b_1 &= 1. ; & b_2 &= 0. & : \text{un zéro} & z &= 0 \\ a_1 &= -1.4 ; & a_2 &= 0.49 & : \text{deux pôles} & z &= 0.7 \end{aligned}$$

Indiquons encore que les essais, dont on précisera les résultats un peu plus loin, ont été effectués sur la base d'un enregistrement d'entrée/sortie de 200 points, avec un rapport bruit sur signal de 10 % environ, mesuré par le rapport des écarts type de b_t et y_t . Deux cas ont été considérés :

- Pb_3 ; $\alpha = 0$.
- Pb_4 ; $\alpha = 0.96$

C'est l'ensemble des résultats numériques associés à ces quatre problèmes que nous abordons maintenant.

3.1 - Etude comparative de différentes techniques de minimisation de l'erreur de sortie

On a comparé cinq méthodes différentes d'optimisation non linéaire, appliquées respectivement au critère J et au critère projeté $\tilde{J}$; ce qui représente globalement dix approches distinctes pour minimiser l'erreur de sortie relative à un modèle donné.

Ces différentes méthodes d'optimisation sont l'algorithme de HOOKE et JEEVES [11] (HJ) bien connu pour sa simplicité, l'algorithme des gradients conjugués dans la version donnée par FLETCHER en 1972 [40] (CG), l'algorithme de FLETCHER [15] du type quasi-Newton, dans sa forme "gradient analytique" (FQN) et dans sa forme "gradients par différences" (FQND), enfin un algorithme de régression non linéaire du type Marquardt dans la version de FLETCHER [17] (FM).

Une première série d'essais a été effectuée sur le premier test - Pb_1 - Dans les tableaux qui suivent NF désigne le nombre de calculs du critère (+ gradients) ou du résidu et du Jacobien, NT représente le nombre d'itérations, enfin CPU (Central Processor Unit) est le temps calcul sur l'unité centrale exprimé ici 1/10000 d'heure, soit 0,36 seconde, et relatif à un IBM 370-135. Précisons enfin que dans tous les cas, on a admis que la convergence était atteinte lorsque les variations sur les paramètres étaient inférieures à 10^{-10} , en valeur relative, la précision machine étant 10^{-16} (double précision).

En ce qui concerne l'initialisation des algorithmes deux techniques ont été retenues :

- La première consiste à donner des valeurs vraisemblables aux paramètres en se basant sur la procédure suivante. Raisonnons sur un modèle d'ordre 2. On choisit alors un pôle double correspondant à un temps de réponse voisin

de celui du système, ce qui est facile à réaliser. On en déduit (à la main !) les coefficients du dénominateur, puis ceux du numérateur tels que les deux premiers points (non nuls) de la séquence estimée soient identiques aux données. Par "Pb₁" on a choisi un pôle double $z = 0,8$ ce qui donne $a_1 = -1.6$; $a_2 = 0.64$; par ailleurs on a $h_1 = 0.6$ et $h_2 = 1.4$, d'où $b_1 = h_1 = 0.6$ et $b_2 = h_2 + a_1 h_1 = 0.44$. La même opération peut bien sûr s'effectuer pour des ordres supérieurs. Dans le cas présent où on a modélisé successivement "Pb₁" par des systèmes d'ordre croissant, on a trouvé plus simple de prendre comme initialisation pour le modèle d'ordre $n+1$ les paramètres du modèle optimum d'ordre n en posant $a_{n+1} = b_{n+1} = 0$.

- La seconde approche réalise une procédure d'initialisation automatique qui consiste à déterminer successivement les coefficients du dénominateur puis ceux du numérateur en résolvant un système linéaire au sens des moindres carrés.

En se référant aux développements du § 2.3.3 et à l'équation (V-58), on en déduit immédiatement que les paramètres a_i d'un modèle d'ordre n , noté $\underline{a}$, sont solution, au sens des moindres carrés, du système :

$$F_{nN} \underline{a} = \underline{f}_{n+1} \quad (V-76)$$

Désignons maintenant par $\underline{b}$ le vecteur de composantes b_i les coefficients du numérateur. La séquence de pondération du modèle étant linéaire par rapport à $\underline{b}$, on peut construire une matrice A ne dépendant que de $\underline{a}$ telle que :

$$A \underline{b} = \underline{f}_m \quad (V-77)$$

où $\underline{f}_m$ est un vecteur contenant la séquence de pondération du modèle de paramètres $\underline{a}$, $\underline{b}$. L'initialisation de $\underline{b}$ est alors obtenue en résolvant, au sens des moindres carrés le système :

$$A \underline{b} = \underline{f} \quad (V-78)$$

où cette fois $\underline{f}$ désigne la séquence de pondération du système donné.

En opérant ainsi, on a obtenu les résultats suivants :

$$- a) \text{ "Pb}_1\text{ " Modèle d'ordre 2 : } \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

Initialisation : $b_0 = 0$; $b_1 = 0.6$; $b_2 = 0.44$; $a_1 = -1.6$; $a_2 = 0.64$

$$J = 0.119947 \cdot 10^2 ; \bar{J} = 0.659505 \cdot 10^1$$

	HJ/J	HJ/J	GC/J	GC/J	FQND/J	FQND/J	FQN/J	FQN/J	FM/J	FM/J
NF	8192	913	619	37	192	69	24	12	18	16
NT	3987	549	236	14	20	13	16	11	15	15
CPU	794	439	366	46	25	37	19	19	31	36

Tableau 2

Initialisation automatique :

$$b_0 = 0.221626.10^{-14} ; b_1 = 0.615005 ; b_2 = 0,279499 ; a_1 = -1.82826 ; a_2 = 0.871737$$

$$J = \tilde{J} = 0.600338.10^1$$

	HJ/J	HJ/J	GC/J	GC/J	FQND/J	FQND/J	FQN/J	FQN/J	FM/J	FM/J
NF	3766	645	369	28	221	72	23	12	19	16
NT	1763	283	136	11	24	14	20	11	15	13
CPU	359	295	214	36	27	40	19	19	28	35

Tableau 3

Résultats communs :

$$b_0 = -0.578351.10^{-7} ; b_1 = 0.303352 ; b_2 = 1.02689 ; a_1 = -1.71421 ; a_2 = 0.762623$$

$$J = \tilde{J} = 0.43768288... \quad (V-79)$$

-b) "Pb₁" Modèle d'ordre 3 :

$$\frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}}$$

Initialisation : les paramètres optimaux précédents tronqués à 3 décimales et
 $b_3 = a_3 = 0.$

$$J = 0.627472 ; \tilde{J} = 0.232062 \quad (V-80)$$

	HJ/J	HJ/J	GC/J	GC/J	FQND/J	FQND/J	FQN/J	FQN/J	FM/J	FM/J
NF	>100*.000	22500	2763	1918	310	122	31	27	33	28
NT	40611	13900	996	765	27	18	28	14	27	25
CPU	10837	>15000	2475	3246	41	86	34	52	61	106

Tableau 4

$$* J = 0.2753.10^{-1} ; ** \tilde{J} = 0.3315.10^{-1} ; (\text{au lieu de } J = \tilde{J} = 0.2601.10^{-1})$$

Initialisation automatique :

$b_0 = 0.409083.10^{-16}$; $b_1 = 0.599357$; $b_2 = 0.106774$; $b_3 = 0.115386$
 $a_1 = -2.15916$; $a_2 = 1.53902$; -0.350456
 $J = \tilde{J} = 0.104557$

	HJ/J	HJ/ $\tilde{J}$	GC/J	GC/ $\tilde{J}$	FQND/J	FQND/ $\tilde{J}$	FQN/J	FQN/ $\tilde{J}$	FM/J	FM/ $\tilde{J}$
NE			3468	782	277	117	28	21	29	28
NT			1259	313	23	18	25	14	22	24
CPU			3008	1320	37	88	30	42	56	106

Tableau 5

Résultats communs, à l'exception de la méthode de HOOKE et JEEVES abandonnée en raison de sa lenteur à converger:

$b_0 = -0.453019.10^{-12}$; $b_1 = 0.558165$; $b_2 = 0.0936807$; $b_3 = -0.214554$;
 $a_1 = -2.45915$; $a_2 = 2.05886$; $a_3 = -0.584592$
 $J = \tilde{J} = 0.2601204 \dots 10^{-1}$
 $-c) \text{ "Pb}_1\text{ " Mod\`ele d'ordre 4 (exact)}$

$$\frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + b_4 z^{-4}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4}} \quad (V-61)$$

Initialisation : les param\`etres optimaux pr\`ec\`edents tronqu\`es \`a 4 d\`ecimales et

$b_4 = a_4 = 0.$
 $J = 0.3468.10^{-1}$; $\tilde{J} = 0.2654.10^{-1}$ (V-62)

	GC/J	GC/ $\tilde{J}$	FQND/J	FQND/ $\tilde{J}$	FQN/J	FQN/ $\tilde{J}$	FM/J	FM/ $\tilde{J}$
NE	9182	4524	542	181	44	31	12	8
NT	3104	1491	34	21	38	23	9	7
CPU	11255	10474	86	180	63	80	39	62
$\tilde{J}$	$0.26.10^{-4}$	$0.10.10^{-8}$	$0.47.10^{-9}$	$0.47.10^{-13}$	$0.16.10^{-24}$	$0.45.10^{-22}$	$0.15.10^{-26}$	$0.28.10^{-26}$

Tableau 6

On a résumé ci-après, en regard des paramètres exacts les valeurs obtenues respectivement dans le moins bon et le meilleur des cas (GC/J et FM/J).

exacts	GC/J	FM/J
$b_0 = 0.$	$0.236.10^{-4}$	$-0.1425.10^{-24}$
$b_1 = 0.6$	0.6007	0.5999999999999999
$b_2 = -0.4$	-0.3914	-0.40000000000000040
$b_3 = 0.015$	0.0235	0.01499999999999911
$b_4 = -0.109$	-0.1203	-1.089999999999980
$a_1 = -3$	-2.977	-3.0000000000000007
$a_2 = 3.33$	3.272	3.3300000000000018
$a_3 = 1.6$	-1.549	-1.6000000000000016
$a_4 = 0.2736$	0.2588	0.2736000000000047

Tableau 7

Remarque : L'initialisation automatique conduisant directement à l'optimum n'a donc pas été appliquée ici.

Cette première série d'essais permet déjà de tirer quelques conclusions.

De toute évidence la méthode de HOOKE et JEEVES n'est pas adaptée à ce type de problème. En effet, par construction, sa convergence est d'autant plus lente que la surface (engendrée par $||\underline{\varepsilon}(\alpha)||$ présente des vallées plus profondes et plus étroites (gradients variant très vite). Or dans le cas présent, ce mauvais conditionnement de la surface croît très rapidement avec le degré du dénominateur. Pour s'en rendre compte, il suffit d'observer les variations du critère lorsqu'on remplace les paramètres à l'optimum par leur valeur tronquée à 3 ou 4 décimales (cf V-79 à V-82). Quant aux gradients, ils peuvent varier pour des conditions analogues dans un rapport atteignant 10^{10} .

En ce qui concerne l'utilisation du critère projeté, dans tous les cas, il a conduit à une diminution du nombre de calculs du critère. Cependant sur le plan temps de calculs, cette diminution ne compense pas nécessairement le coût supplémentaire par évaluation du critère.

Dans ce domaine, il est intéressant de noter que la méthode de régression non linéaire est généralement moins performante, sauf lorsque le minimum est nul[†], condition qui ne sera pas a priori remplie en pratique. Par ailleurs, le peu de différence du nombre de calculs du critère que l'on a enregistré entre les deux solutions FM/J et FM/ $\tilde{J}$ s'explique par la façon dont sont construits les itérés successifs (linéarisation de l'écart $\underline{g}$ autour du point courant).

Ces diverses constatations nous ont amené à ne conserver par la suite que la méthode de FLETCHER FQN (FQND étant moins rapide, cf Tableaux 2 à 6) et, pour comparaison, l'algorithme de régression non linéaire FM, étant seul, ici, à tenir compte de la forme du critère en somme de carrés.

On a résumé au Tableau 8 l'ensemble des résultats obtenus avec le deuxième test "Pb₂", directement défini par une séquence de pondération (réponse d'un surchauffeur).

Nous remarquerons d'abord que la tendance enregistrée précédemment ne se conserve que pour les modèles d'ordre 2 et 3. Au delà la situation change progressivement du tout au tout au fur et à mesure que la complexité du problème à résoudre augmente. La multiplicité des minima est un point nouveau qui rend d'ailleurs plus délicat la comparaison des différents résultats. On a constaté ici et sur d'autres essais de ce type que les minima relatifs les plus bas étaient les plus difficiles à atteindre en raison de l'extrême sensibilité du critère dans leur voisinage. On a, par exemple, noté les résultats suivants, $\underline{g}$ désignant le gradient de J ou de $\tilde{J}$.

$$\text{FQN/J ordre 5 : } J = 0.7975.10^1, \quad ||\underline{g}|| = 10^{-5}$$

$$\text{FQN}/\tilde{J} \text{ ordre 5 : } \tilde{J} = 0.7918.10^1, \quad ||\underline{g}|| = 10^2$$

$$\text{FQN/J ordre 6 : } J = 0.1462.10^1, \quad ||\underline{g}|| = 10^{12}$$

$$\text{FQN}/\tilde{J} \text{ ordre 6 : } \tilde{J} = 0.7613.10^1, \quad ||\underline{g}|| = 10^{-4}$$

Sans prétendre tirer de conclusions, on peut souligner que les gradients associés aux divers minima sont d'autant plus grands, que le minimum correspondant est petit.

[†] Lorsque le minimum est nul, de linéaire la convergence devient sur-linéaire voire quadratique !

		FQN/J	FQN/ $\tilde{J}$	FM/J	FM/ $\tilde{J}$	
deux minima	NF	17	12	53	48	ordre 2
	NT	15	11	49	43	
	J/ $\tilde{J}$	$0.6101 \cdot 10^2$	$0.6101 \cdot 10^2$	$0.6101 \cdot 10^2$	$0.6101 \cdot 10^2$	
	CPU	17	20	71	90	
	NF	37	24	21	16	ordre 3
	NT	26	20	15	15	
	J/ $\tilde{J}$	$0.8006 \cdot 10^1$	$0.8006 \cdot 10^1$	$0.8006 \cdot 10^1$	$0.8006 \cdot 10^1$	
	CPU	42	48	41	64	
	NF	35	27	20	25	ordre 4
	NT	29	18	14	9	
	J/ $\tilde{J}$	$0.7992 \cdot 10^1$	$0.7999 \cdot 10^1$	$0.7992 \cdot 10^1$	$0.7892 \cdot 10^1$	
	CPU	55	70	55	160	
quatre minima	NF	43	16	63	27	ordre 5
	NT	33	7	48	10	
	J/ $\tilde{J}$	$0.7975 \cdot 10^1$	$0.7918 \cdot 10^1$	$0.1262 \cdot 10^1$	$0.738 \cdot 10^1$	
	CPU	89	57	227	273	
quatre minima	NF	2066	38	1200	23	ordre 6
	NT	1394	29	202	7	
	J/ $\tilde{J}$	$0.1462 \cdot 10^1$	$0.7613 \cdot 10^1$	$0.1248 \cdot 10^1$	$0.775 \cdot 10^1$	
	CPU	4680	156	11929	361	

Tableau 8

Au terme de cette brève étude expérimentale, nous constatons après bien des auteurs que ce type d'analyse comparative ne permet généralement pas de définir de tendances très significatives, et c'est une des raisons pour lesquelles nous ne lui avons pas donné plus d'ampleur dans le cadre de ce travail.

Nous proposons néanmoins les quelques suggestions suivantes :

- Les méthodes heuristiques telle que HOOKE et JEEVES sont inapplicables à ce type de problème d'optimisation dont la sensibilité par rapport à certains paramètres croît considérablement avec l'ordre du système (cf. Tableau 4).

- Nous pensons que les techniques qui prennent en compte la structure "somme de carrés" du critère ne jouissent pas de performance justifiant *a priori* leur utilisation ; nous préconisons plutôt d'aborder ce problème de modélisation en utilisant une méthode du type quasi-Newton (cf CPU du Tableau 8).

- D'un autre côté, l'utilisation du critère projeté, sans être une solution de référence, est un moyen efficace de traiter des problèmes difficiles à résoudre, et que l'approche classique ne permet pas d'appréhender dans des temps de calculs raisonnables (cf Tableau 8, ordre 6 où l'on observe un gain de temps dans le rapport 1 à 30 !).

Par la suite les résultats présentés dans le cadre de l'approximation des systèmes mono ou multivariables, à partir d'entrées/sorties quelconque ont été obtenus en faisant appel à la méthode de FLETCHER [15] du type quasi Newton. Cependant cette technique n'a sans doute pas été exploitée dans les meilleures conditions compte tenu de l'analyse précédente, réalisée chronologiquement après la mise en place des programmes généraux d'approximation multivariables.

3.2 - Identification d'un système simulé

Notre propos est d'illustrer à l'aide d'un exemple simple (système du 2ème ordre) la méthode numérique développée § 2. Bien que les conclusions qu'on puisse tirer de ce genre d'exemple soient déjà connues, GUSTAVSON [28], [30], FOULARD [41] [29], il nous paraissait impossible de ne pas inclure de tels résultats dans ce mémoire. Par ailleurs ces essais, bien qu'un peu académique mettent néanmoins en lumière la démarche que nous suivrons dans le cas multivariable, beaucoup plus complexe.

a) " Pb_3 " excitation blanche, système du 2ème ordre, perturbations indépendantes.

On a calculé des réalisations d'ordre 1 à 4 obtenues suivant les deux approches fondamentales : minimisation de l'erreur d'équation, minimisation de l'erreur de sortie.

Les caractéristiques de ces huit modèles ont été résumées au Tableau 9 et à la Figure 12, avec les notations suivantes :

- $||\underline{e}^*||$ erreur d'équation, modèle obtenu par minimisation de l'erreur d'équation.
- $||\underline{\varepsilon}^*||$ erreur de sortie, modèle obtenu par minimisation de l'erreur d'équation.
- $||\hat{\underline{\varepsilon}}||$ erreur de sortie, modèle obtenu par minimisation de l'erreur de sortie

Ces grandeurs exprimées en % sont en fait des erreurs relatives obtenues en divisant la norme $||\underline{e}||$ ou $||\underline{\varepsilon}||$ par $\sum_t y_t^2$ où y_t est la sortie mesurée.

ordre	1	2	3	4
$ \underline{e}^* $	24,7 %	18 %	15,5 %	13,2 %
$ \underline{\varepsilon}^* $	48,9 %	21,9 %	14,27 %	11,99 %
$ \hat{\underline{\varepsilon}} $	33 %	11,21 %	11,14 %	10,98 %

Tableau 9

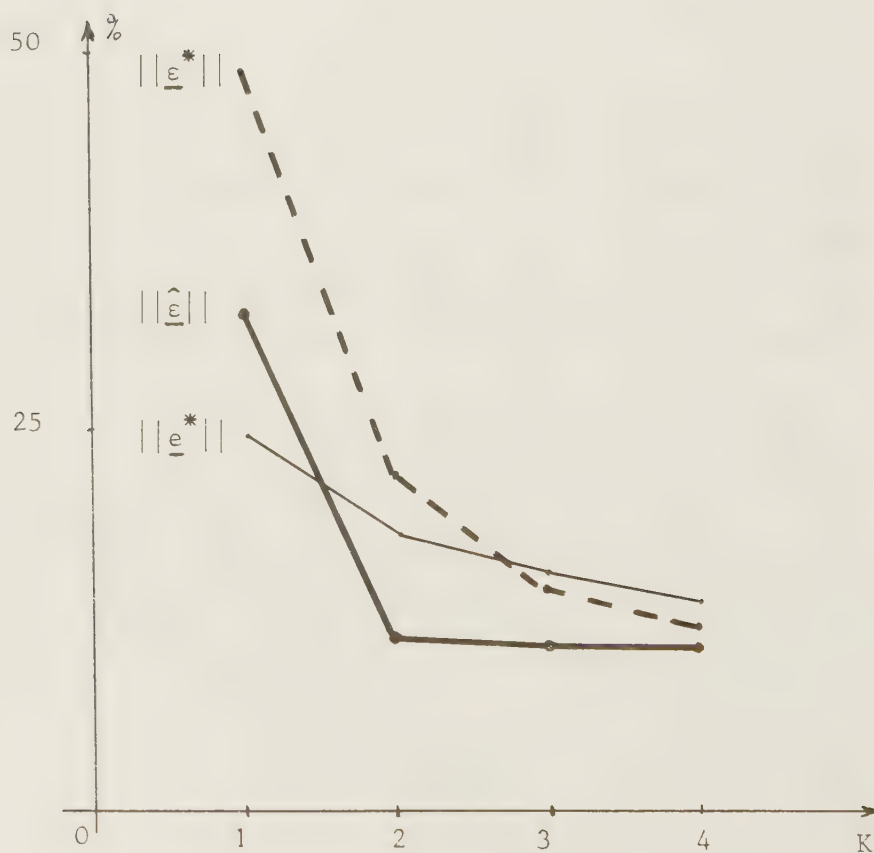


Figure 12

modèle "erreur d'équation"		modèle "erreur de sortie"	
Zéros			
Poles	0.9411	0.8808	Ordre 1
Modèle	$\frac{0.9657 Z^{-1}}{1-0.9411 Z^{-1}}$	$\frac{1.4142 Z^{-1}}{1-0.8808 Z^{-1}}$	
Zéros	-0.28890	<u>Valeurs exactes</u>	Ordre 2
Poles	0.18566 ; 0.87401	Zéro : 0. Poles : 0.7 ; 0.	
Modèle	$\frac{0.8997 Z^{-1} + 0.2859 Z^{-2}}{1-1.0596 Z^{-1} + 0.1622 Z^{-2}}$	$\frac{1 Z^{-1} + 0. Z^{-2}}{1-1.4 Z^{-1} + 0.49 Z^{-2}}$ système :	
Zéros	- 0.31147 ± i 0.23381	0.28414 10 ⁻¹ : - 0.59441	Ordre 3
Poles	0.39412 ; 0.83215 ; -0.50837	0.71132 ; 0.70734 ; -0.64527	
Modèle	$\frac{1.016 Z^{-1} + 0.6329 Z^{-2} + 0.1541 Z^{-3}}{1-0.7175 Z^{-1} - 0.2958 Z^{-2} + 0.1668 Z^{-3}}$	$\frac{1.010 Z^{-1} + 0.5717 Z^{-2} - 0.017 Z^{-3}}{1-0.7763 Z^{-1} - 0.4093 Z^{-2} + 0.3279 Z^{-3}}$	
Zéro	- 0.31094 ; - 0.30606 ± i 0.56357	- 0.1040 10 ⁻² ; - 0.73378 ± i 0.57961	Ordre 4
Poles	- 0.42493 ± i 0.55517 ; 0.50413 ; 0.79837	0.70625 + i 0.0380 ; - 0.74887 ± i 0.59563	
Modèle	$\frac{0.9941 Z^{-1} + 0.9176 Z^{-2} + 0.5980 Z^{-3} + 0.1271 Z^{-4}}{1-0.4526 Z^{-1} - 0.2156 Z^{-2} - 0.2945 Z^{-3} + 0.1966 Z^{-4}}$	$\frac{0.9841 Z^{-1} + 1.445 Z^{-2} + 0.8620 Z^{-3} + 0.0009 Z^{-4}}{1+0.0852 Z^{-1} - 0.6996 Z^{-2} - 0.5439 Z^{-3} + 0.4580 Z^{-4}}$	

T A B L E A U 10

Pour compléter ces informations, on a calculé les pôles et les zéros de ces différents modèles. Ceux-ci ont été regroupés au Tableau 10, avec les transmittances correspondantes, de manière à faire apparaître les simplifications (approchées) qu'on est en droit d'attendre.

Il est clair, ici, que l'ordre du système ($n=2$) ne peut être déterminé sans ambiguïté, sans faire appel à la minimisation de l'erreur de sortie. C'est en effet dans ce cas là que la courbe d'un des critères d'erreur (erreur de sortie minimisée) présente un coude franc (pour $n = 2$) et au-delà duquel pôles et zéros surabondants se simplifient avec une bonne approximation. Malgré le caractère un peu simpliste de cet exemple, ces observations sont assez générales. L'expérience confirme, en effet, qu'en présence de perturbations peu corrélées, l'approche "erreur d'équation" ne permet pas de construire des approximations de dimension réduite satisfaisantes. Il faut en effet atteindre l'ordre 4 pour obtenir avec cette technique une réponse indicielle proche de celle du système original (Figure 13). Cependant en tant que modèle de conduite (destiné à effectuer la synthèse d'un régulateur) une approximation par erreur d'équation de faible dimension peut néanmoins être largement suffisant, tout en présentant une erreur non négligeable (en boucle ouverte) Fig. 13.

Remarques : En revenant à la Fig. 12, on notera la convergence de $||\underline{\varepsilon}^*||$, $||\hat{\underline{\varepsilon}}||$ et $||\underline{e}^*||$ vers une même valeur, lorsque l'ordre du modèle augmente. Il ne s'agit pas là d'un hasard ; cette propriété est due à la nature indépendante des perturbations. On constatera d'ailleurs, Fig. 14, que dans le cas de perturbations corrélées, l'erreur de sortie $||\underline{\varepsilon}^*||$ et $||\hat{\underline{\varepsilon}}||$ convergent vers une même valeur mais différente de celle de l'erreur d'équation $||\underline{e}^*||$. Nous ne donnerons pas ici de démonstration de ce point, qui ne présente pas un intérêt particulier dans le cas présent.

b) "Pb₄" perturbations fortement corrélées

On a effectué de nouveau la même série d'essais qui ont conduit cette fois aux valeurs suivantes : Tableau 11 et Figure 14 en ce qui concerne les critères d'erreurs, et Tableau 12, en ce qui concerne les transmittances zéros et pôles des modèles.

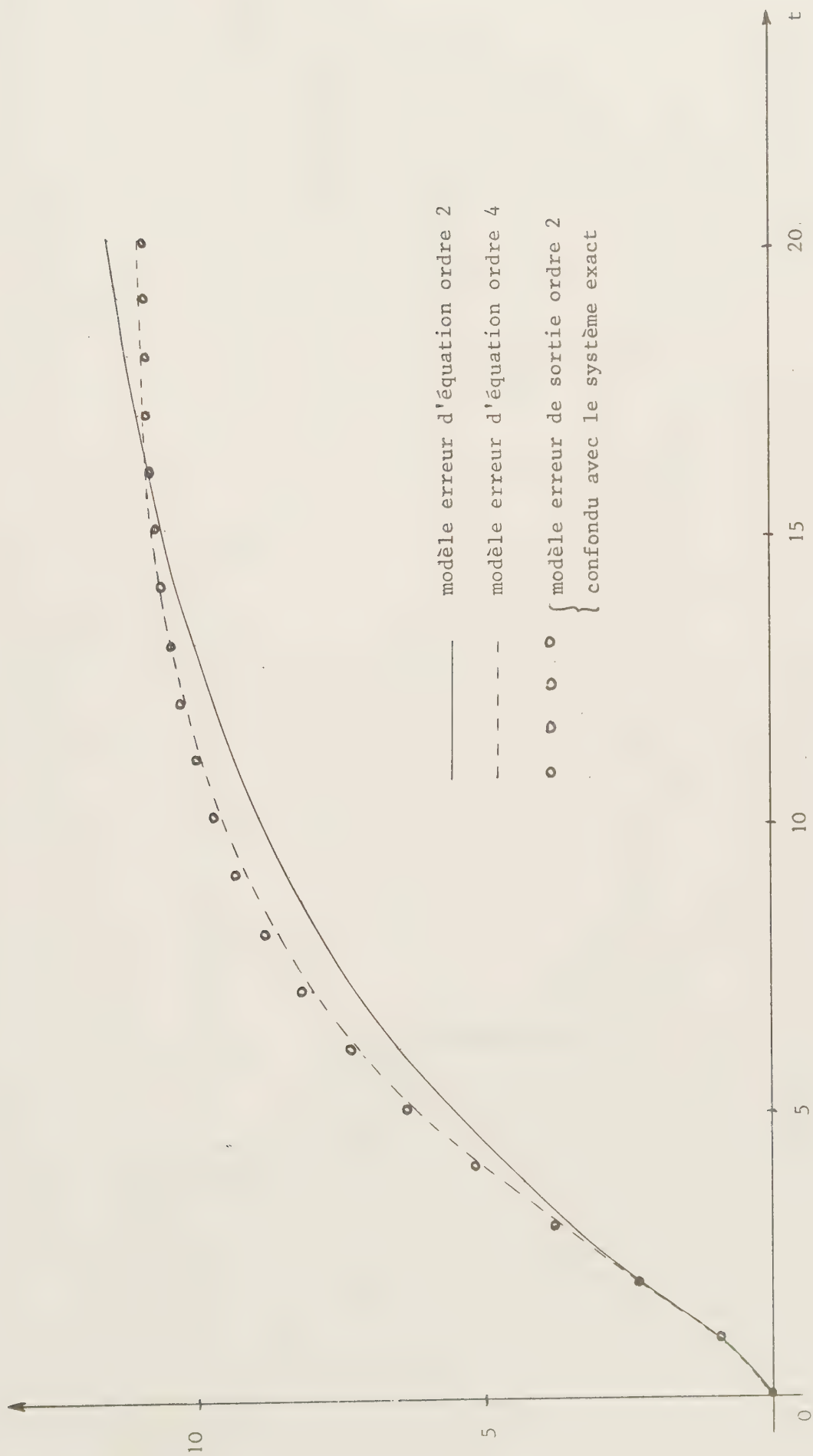


Figure 13

ordre	1	2	3	4
$ \underline{e}^* $	20 %	3,47 %	3,29 %	3,2 %
$ \varepsilon^* $	58,7 %	9,71 %	9,53 %	9,52 %
$ \hat{\varepsilon} $	32,9 %	9,45 %	9,43 %	9,44 %

Tableau 11

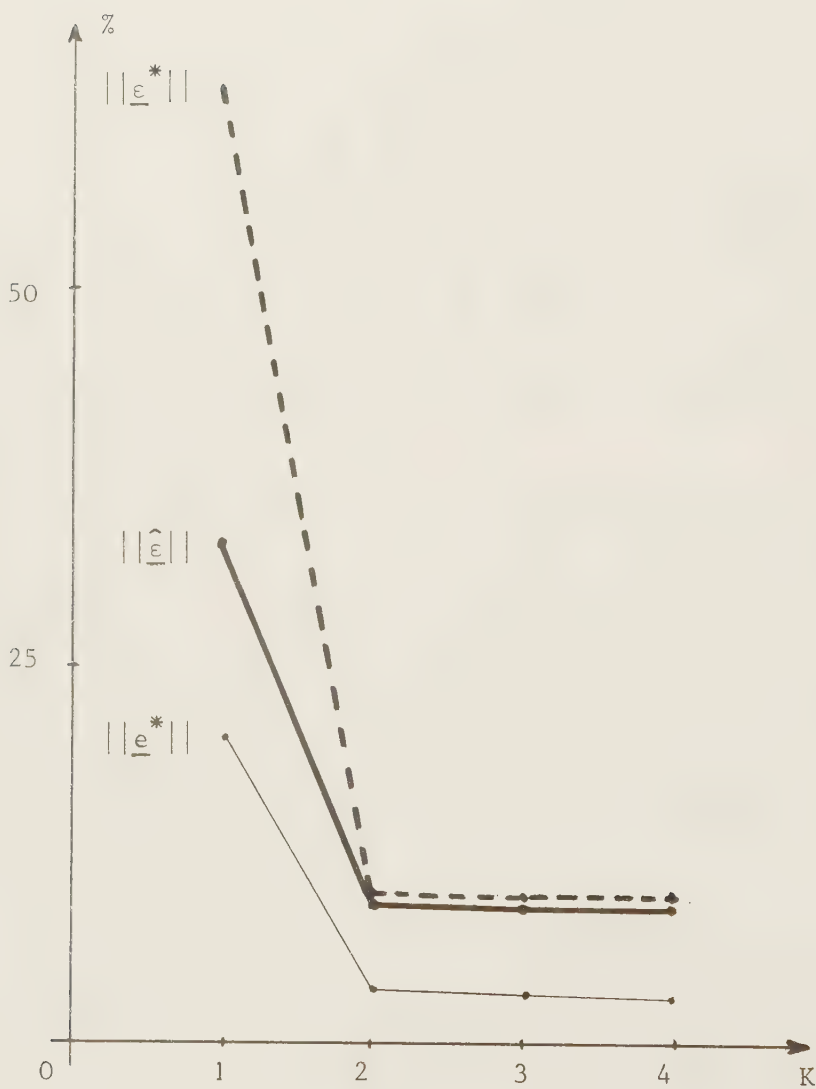


Figure 14

modèle "erreur d'équation"		modèle "erreur de sortie"	
Zéros			
Pôles	0.9516	0.8771	Ordre 1
Modèle	$\frac{0.9660 Z^{-1}}{1-0.9516 Z^{-1}}$	$\frac{1.4440 Z^{-1}}{1-0.8771 Z^{-1}}$	
Zéros	$-0.3616 \cdot 10^{-2}$	<u>Valeurs exactes</u> Zéro : 0. Pôles : 0.7 ; 0.7 système : $\frac{1 \cdot Z^{-1} + 0 \cdot Z^{-2}}{1-1.4Z^{-1} + 0.49Z^{-2}}$	Ordre 2
Pôles	0.63454 ; 0.75247		$\frac{0.5456 \cdot 10^{-1}}{0.70978 \pm i 0.0635}$
Modèle	$\frac{1.003Z^{-1} + 0.00361Z^{-2}}{1-1.3870Z^{-1} + 0.4775Z^{-2}}$	$\frac{1.026Z^{-1} - 0.0560 Z^{-2}}{1.-1.4195Z^{-1} + 0.5078Z^{-2}}$	
Zéros	$0.2504 \cdot 10^{-1}$; -0.31377	0.11397 ; -0.41106 0.71382 ; 0.71316 ; - 0.32670	Ordre 3
Pôles	0.70238 ; 0.70008 ; -0.32787		
Modèle	$\frac{1.005Z^{-1} + 0.3043Z^{-2} - 0.0082Z^{-3}}{1-1.0887Z^{-1} + 0.0518Z^{-2} + 0.1543Z^{-3}}$	$\frac{1.0130Z^{-1} + 0.3009Z^{-2} - 0.0474Z^{-3}}{1-1.1003Z^{-1} + 0.0477Z^{-2} + 0.1679Z^{-3}}$	
Zéros	$-0.3966 \cdot 10^{-1}$; -0.16207 ± i 0.37878	-0.4783.10 ⁻¹ ; -0.15797 ± i 0.37787 0.70247 ± i 0.00511 ; -0.19394 ± i 0.39921	Ordre 4
Pôles	0.65812 ; 0.73397 ; -0.18221 ± i 0.38312		
Modèle	$\frac{1.0028Z^{-1} + 0.3648Z^{-2} + 0.1831Z^{-3} + 0.0067Z^{-4}}{1-1.0276Z^{-1} + 0.1557Z^{-2} - 0.0745Z^{-3} + 0.0869Z^{-4}}$	$\frac{1.0179Z^{-1} + 0.3703Z^{-2} + 0.1861Z^{-3} + 0.0081Z^{-4}}{1-1.0170Z^{-1} + 0.1481Z^{-2} - 0.0843Z^{-3} + 0.0977Z^{-4}}$	

T A B L E A U 12

De ces essais, il ressort clairement que toutes ces informations s'accordent pour indiquer que l'ordre du système est deux. On constate également que pour n supérieur ou égal à 2, la minimisation de l'erreur de sortie n'apporte pratiquement aucune amélioration du critère par rapport à sa valeur initiale, obtenue en minimisant l'erreur d'équation. On notera parallèlement qu'au fur et à mesure que l'ordre du modèle augmente, la procédure "erreur d'équation" conduit à des paramètres (zéros et pôles compris) de plus en plus voisins de ceux du modèle "erreur de sortie". Cette similitude des résultats explique pourquoi nous n'avons pas tracé les réponses indicielles de ces différents modèles qui auraient été plus ou moins confondues avec la réponse exacte du système. On remarquera par ailleurs (Fig. 14) qu'erreur de sortie et erreur d'équation ont "convergé" vers des valeurs différentes (en raison de la corrélation du bruit sur la sortie).

Rappelons qu'enfin l'approche par minimisation de l'erreur de sortie n'est plus ici identique à la méthode du maximum de vraisemblance (perturbations corrélées), et par conséquent ne s'impose plus comme "le" traitement optimal des données (au sens du maximum de vraisemblance).

Il est bien évident qu'en pratique la situation n'est pas aussi claire, cf. § V, et l'utilisation combinée de plusieurs informations demeurent nécessaire.

Remarques : L'intérêt des problèmes de réduction ou/et d'approximation se situant essentiellement au niveau des systèmes multivariables, nous n'avons pas insisté, ici, sur ce point dont une illustration élémentaire est donnée par le modèle du premier ordre.

En conclusion de ces essais réservés au cas monovariable, nous retiendrons les deux points suivants :

- La minimisation de l'erreur de sortie, par programmation non linéaire, peut être considérée comme une technique de référence, à défaut d'une approche statistique.

- La minimisation de l'erreur d'équation est capable de conduire à de très bons résultats dans certains cas et est généralement suffisante pour élaborer un modèle de conduite. [44]

Ces remarques justifient, à notre sens, la position hybride, à laquelle on a déjà fait allusion dans son principe, et qui consiste à "dégrossir" le problème sur la base du critère erreur d'équation, puis de minimiser l'erreur de sortie.

Cette dernière opération, beaucoup plus lourde que la précédente, pourrait être approchée en se limitant aux paramètres intervenant linéairement dans ε_t . Cette solution qui reviendrait en fait à ne calculer qu'une fois le critère projeté $\tilde{J}$ associé au critère global d'erreur de sortie $J = ||\underline{\varepsilon}||^2$, n'est effectivement intéressante que dans le cadre des systèmes multivariables (que nous allons aborder maintenant) en raison du grand nombre de degrés de liberté que représentent alors les matrices Γ et D .

4 - APPROXIMATION DES SYSTEMES MULTIVARIABLES -

Dans le cadre de la réalisation exacte d'un système, une des caractéristiques essentielles des formes canoniques utilisées (couplage dans un seul sens) était de permettre un traitement séquentiel du problème sortie par sortie, cf. Algorithmes 1 à 6.

Une des difficultés fondamentales apportée par le problème d'approximation se résume dans l'impossibilité de réaliser exactement une telle hiérarchisation des calculs en raison même de l'écart qui existe entre modèle et objet. Il est clair, en effet, que la présence des couplages entre sorties rend solidaires les unes des autres les erreurs $||\underline{\varepsilon}||$ associées à chaque sortie (sauf évidemment lorsqu'elles sont nulles : réalisation exacte). Le problème d'approximation demande donc, en toute rigueur, un traitement simultané des sorties, résultat que l'on retrouve implicitement dans l'approche stochastique du maximum de vraisemblance.

Considérons le modèle défini dans la forme innovation (ou forme filtre) :

$$\begin{cases} \underline{x}_{k+1} = \Phi \underline{x}_k + \Gamma \underline{u}_k + K \underline{e}_k \\ \underline{y}_k = H \underline{x}_k + D \underline{u}_k + \underline{e}_k \end{cases} \quad (V-83)$$

où $\{\underline{e}_k\}$ est une séquence indépendante gaussienne. Dans ces conditions, la détermination de l'ensemble des paramètres du système $(\Phi, \Gamma, H, D, K) = (\underline{\theta})$ est obtenue [35] , [36] en minimisant le critère :

$$V(\underline{\theta}) = \det \left[\sum_k \underline{e}'_k \underline{e}'_k{}^T \right] \quad (V-84)$$

avec :

$$\begin{cases} \underline{x}'_{k+1} = \Phi \underline{x}'_k + \Gamma \underline{u}_k + K(\underline{y}_k - H \underline{x}'_k - D \underline{u}_k) \\ \underline{e}'_k = \underline{y}_k - H \underline{x}'_k - D \underline{u}_k \end{cases}$$

expression dans laquelle $\{\underline{u}_k, \underline{y}_k\}$ sont les couples entrées/sorties du processus

objet. Sur la base de cette formulation, la détermination directe d'une forme canonique n'en demeure pas moins difficile. MAYNE suggère de résoudre (V-84) autant de fois qu'il est nécessaire en faisant varier les n_i jusqu'à ce que l'augmentation de $n = \sum n_i$ ne contribue plus à faire diminuer d'une manière significative $V(\hat{\theta})$, (au moyen de test statistique).

Il est bien évident qu'une telle approche représente un coût de calcul considérable dès que les entiers n_i et le nombre d'entrées/sorties dépassent quelques unités. Il est à ce sujet surprenant qu'aucune autre approche constructive n'ait été développée (à notre connaissance), si l'on fait exception de la méthode de MITRA [6] applicable seulement à la réduction de modèle et assez loin de cet objectif (les pôles étant issus d'une sélection de ceux du système original). Par ailleurs on ne peut pas considérer que les articles récents d'APLEVICH [37], [38] et GALIANA [39], développés dans le contexte de la minimisation d'un critère erreur de sortie, apportent quelque éléments réellement utilisables dans la résolution pratique du problème : en effet, leurs principaux résultats se situent au niveau d'un formalisme matriciel concis des conditions nécessaires satisfaites à l'optimum, mais dont la concision même masque la difficulté et la quantité effective de calculs à réaliser. Par ailleurs cette approche ne prend nullement en compte la notion de structure canonique qui pourtant permet de diminuer considérablement la dimension du problème à résoudre.

Ces remarques mettent en fait en évidence la nécessité d'établir un compromis entre l'optimalité d'une solution et le réalisme de sa mise en oeuvre.

Nous allons développer dans ce sens un ensemble de solutions "quasi optimales" basées sur un traitement *séquentiel* des sorties et non *simultané*.

Cette hypothèse de travail est motivée d'une part, par la structure des formes canoniques utilisées et d'autre part, parce qu'elle permettra de s'inspirer des algorithmes de réalisation (exacte) et d'assurer aux calculs une structure hiérarchisée dont le coût reste raisonnable.

En suivant cette ligne directrice, nous allons maintenant reprendre les deux principaux thèmes : minimisation de l'erreur d'équation et de l'erreur de sortie.

4.1- Une approche par erreur d'équation

On admet donc que les sorties ont été classées a priori dans un ordre donné, le problème se ramène alors de nouveau à déterminer l'ensemble des invariants n_i , α_{ijk} et les matrices Γ et D caractérisant les formes canoniques observables développées chapitre III. Sans restreindre la généralité de ce qui suit,

nous considérerons, a priori, la forme canonique (2), les données se présentant sous la forme d'un ensemble d'entrées/sorties quelconques $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, \dots, N$. Notre objectif est donc d'extrapoler l'algorithme 5 suivant une démarche analogue à celle empruntée pour traiter le cas monosortie.

Pour cela reprenons les définitions (IV-91) à (IV-96) en imbriquant colonnes d'entrées et colonnes de sorties de la même manière qu'en (V-3).

Nous introduisons ainsi les notations suivantes relatives à un système multivariable m entrées - p sorties d'ordre $n \leq q$:

$$R_{\ell}^i = \begin{bmatrix} \underline{u}_0^T & y_0^i & & \underline{u}_{\ell}^T & y_{\ell}^i \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \underline{u}_{N'}^T & y_{N'}^i & & \underline{u}_{N'+\ell}^T & y_{N'+\ell}^i \end{bmatrix} \begin{matrix} \updownarrow \\ N'+1 ; i=1, \dots, p \end{matrix} \quad (V-85)$$

$\xleftarrow{\quad (\ell+1) \quad (m+1) \quad}$

puis :

$$Y^i = \begin{bmatrix} y_0^i & \dots & y_{n_i-1}^i \\ \vdots & & \vdots \\ y_{N'}^i & \dots & y_{N'+n_i-1}^i \end{bmatrix} \begin{matrix} \updownarrow \\ N'+1 ; i=1, \dots, p \end{matrix} \quad (V-86)$$

$\xleftarrow{\quad n_i \quad}$

et :

$$\begin{cases} S_{\ell}^i = [Y^1 \mid Y^2 \mid \dots \mid Y^{i-1} \mid R_{\ell}^i] ; i = 2, \dots, p \\ S_{\ell}^1 = R_{\ell}^1 \end{cases} \quad (V-87)$$

... où on a noté $N' = N-q$

On définira enfin les partitions :

$$R_{\ell}^i \triangleq \begin{bmatrix} \tilde{R}_{\ell}^i & \tilde{y}_{\ell}^i \end{bmatrix} \quad (V-88)$$

et

$$S_{\ell}^i \triangleq \begin{bmatrix} \tilde{S}_{\ell}^i & \tilde{y}_{\ell}^i \end{bmatrix} \quad (V-89)$$

$\xleftarrow{\quad v_i \quad 1 \quad}$

avec :

$$\begin{cases} v_i = s_{i-1} + (\ell + 1) m + \ell \\ s_i = \sum_{j=1}^i n_j ; s_0 = 0 \end{cases} \quad (V-90)$$

L'approche par erreur d'équation consiste alors à appliquer l'algorithme 5 en substituant à la précision machine une quantité positive petite ζ_i ($i=1, \dots, p$) définissant un critère d'erreur relative. L'invariant n_i est alors déterminé comme étant la première valeur de ℓ pour laquelle :

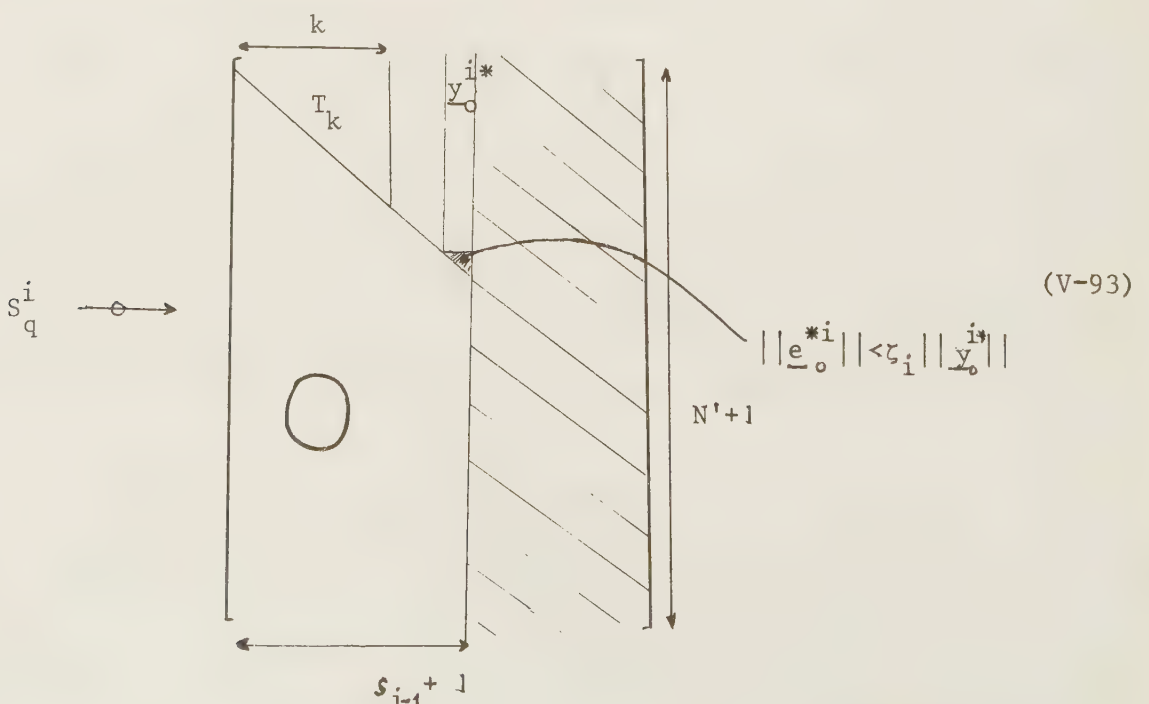
$$||\underline{e}_\ell^{*i}|| \leq \zeta_i ||\underline{y}_\ell^i|| \quad (V-91)$$

avec :

$$||\underline{e}_\ell^{*i}|| = \min_{\underline{\alpha}_i} ||\hat{S}_{\ell-i}^{i\alpha_i} - \underline{y}_\ell^i|| = ||\hat{S}_{\ell-i}^{i\alpha_i^*} - \underline{y}_\ell^i|| \quad (V-92)$$

où $\underline{\alpha}_i$ est l'ensemble des paramètres intervenant dans la modélisation de la $i^{\text{ème}}$ sortie (Cf aussi IV-58).

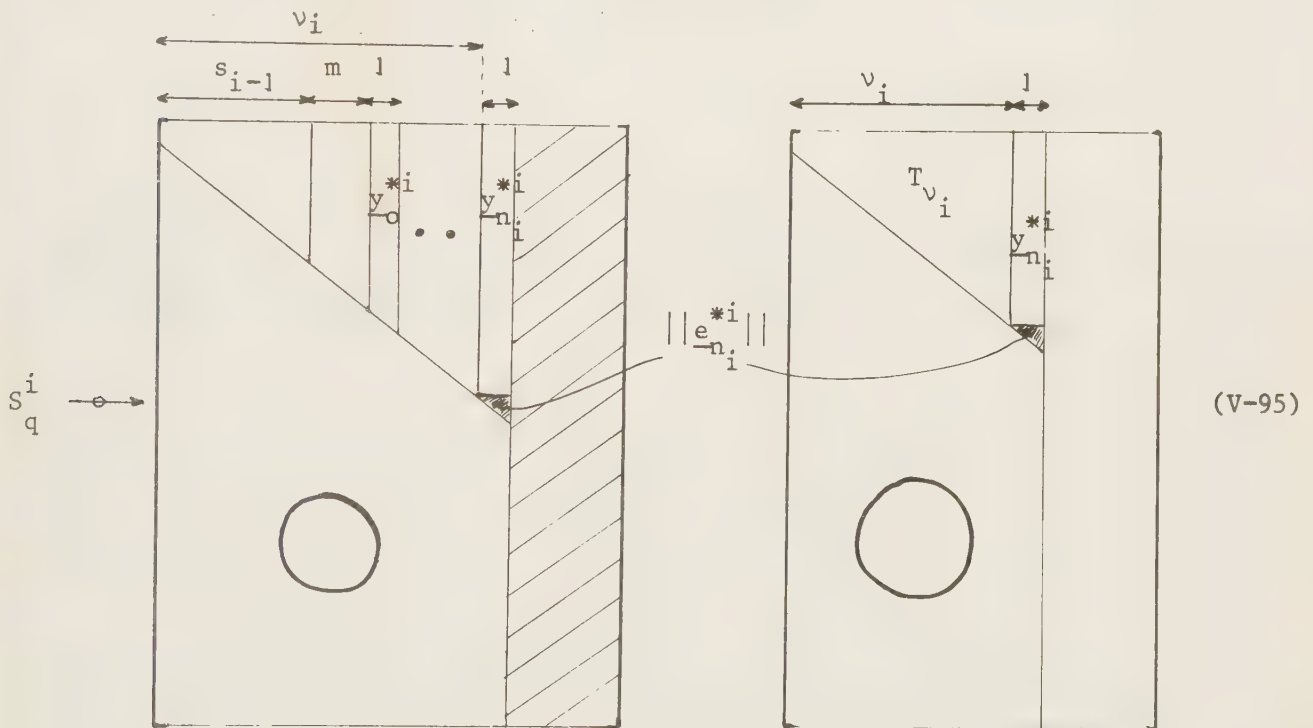
Là encore l'utilisation des transformations de Householder permet de déterminer séparément n_i et $\underline{\alpha}_i^*$ en raisonnant toujours sur la matrice S_q^i transformée (triangularisée). Deux cas sont à envisager suivant que la modélisation de la $i^{\text{ème}}$ sortie nécessite ou non l'introduction de nouveaux états. Considérons d'abord l'hypothèse de la transformation (V-93).



où $k < s_{i-1}$. Ceci signifie que la modélisation approchée de la $i^{\text{ème}}$ sortie ne nécessite aucun état supplémentaire. On en déduit $n_i = 0$ et les s_{i-1} paramètres (invariants) relatifs à cette sortie sont donnés par :

$$\underline{\alpha}_i^* = \begin{bmatrix} T_k^{-1} y_o^{i*} \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \uparrow k \\ \downarrow s_{i-1}-k \end{matrix} \quad (V-94)$$

Plaçons nous maintenant dans la seconde hypothèse $k > s_{i-1}$ ($n_i > 0$). L'invariant n_i est alors facilement obtenu en observant la valeur de l pour laquelle le terme diagonale relatif à une colonne de sortie de la matrice (triangularisée) satisfait (en module) l'inégalité (V-91).



On en déduit immédiatement les paramètres relatifs à la sortie i , soit :

$$\underline{\alpha}_i^* = T_{v_i}^{-1} y_{n_i}^{i*} \quad (V-96)$$

vérifiant (V-92)

Ces opérations étant effectuées pour $i = 1, \dots, p$; la matrice Γ est de nouveau obtenue en résolvant le système triangulaire (IV-57). D'où :

Algo. Tome 5 bis

1 - Etant donné la suite $\{\underline{u}_t, \underline{y}_t\} \ t = 0, \dots, N$; permettant d'identifier un système d'ordre inférieur ou égal à q (donné) ; poser $N' = N - q$.

2 - Pour $i = 1, \dots, p$ construire S_q^i et déterminer n_i et $\underline{\alpha}_i^*$ d'après (V-93)-(V-94) ou (V-95)-(V-96) suivant que n_i est nul ou positif.

Effectuer les partitions suivantes :

$$n_i > 0: (\underline{\alpha}_i^*)^T = [\alpha_{i11} \dots \alpha_{in_i 1} \dots \alpha_{i, i-1, 1} \dots \alpha_{i, i-1, n_{i-1}} \mid \beta_{i1}^T, \alpha_{i11}, \dots, \beta_{in_i}^T, \alpha_{iin_i} \mid \underline{d}_i^T] \quad (V-98)$$

$$n_i = 0: (\underline{\alpha}_i^*)^T = [\alpha_{i11} \dots \alpha_{in_i 1} \dots \alpha_{i, i-1, 1} \dots \alpha_{i, i-1, n_{i-1}}] \quad (V-99)$$

3 - Résoudre le système triangulaire (IV-57) par rapport aux $\underline{y}_{ik}$

4 - Construire l'approximation (Φ, Γ, H, D) suivant la forme définie en (IV-42) et (IV-43), c'est-à-dire :

$$\Phi = \begin{bmatrix} \phi_{11} & & & \\ \phi_{21} & \phi_{22} & & \\ \vdots & & \ddots & \\ \phi_{p1} & \dots & \dots & \phi_{pp} \end{bmatrix} \text{ avec } \phi_{ij} = \begin{bmatrix} \bigcirc \\ \alpha_{ij1} \dots \alpha_{ijn_j} \end{bmatrix} \ i > j \text{ et } \phi_{ii} = \begin{bmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ \alpha_{ii1} & \dots & \alpha_{iin_i} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_p \end{bmatrix} \text{ avec } \Gamma_i = \begin{bmatrix} \underline{y}_{i1}^T \\ \vdots \\ \underline{y}_{in_i}^T \end{bmatrix} ; \ D = \begin{bmatrix} \underline{d}_1^T \\ \vdots \\ \underline{d}_p^T \end{bmatrix}$$

$$H = \begin{bmatrix} \underline{h}_1^T \\ \vdots \\ \underline{h}_p^T \end{bmatrix} \text{ avec } \underline{h}_i^T = \begin{cases} [0 \dots 0 \mid 1 \mid 0 \dots 0] & \text{si } n_i > 0 \\ [\alpha_i^{+T} \mid 0 \dots 0] & \text{si } n_i = 0 \end{cases}$$

$\xleftarrow{s_{i-1}} \quad \xrightarrow{s_{i-1}}$

5 - Calculer l'état initial $\underline{x}_0$ en utilisant successivement l'équation (IV-53) pour $i = 1, \dots, p$, en prenant $t = 0$.

Remarque : On pourrait envisager des algorithmes "i bis", $i = 1, \dots, 6$ dans des termes analogues à celui-ci en s'inspirant de la démarche des algorithmes "i" correspondant, les paramètres étant dans chacun des cas estimés au sens des moindres carrés.

4.2 - Minimisation de l'erreur de sortie

Elle se présente dans des conditions semblables au cas monovariable en raison de la forme des équations (IV-38). Désignons par η_t^i la sortie modèle numéro i à l'instant t . Pour une structure donnée (n_i) on en déduit immédiatement les relations :

$$\eta_{t+n_1}^1 = \sum_{k=1}^{n_1} (\alpha_{11k} \eta_{t+k-1}^1 + \beta_{1k}^T u_{t+k-1}) + d_{1-t+n_1}^T u_{t+n_1} \quad (V-99)$$

avec des conditions initiales $(\underline{\eta}_0^1)^T = (\eta_0^1, \dots, \eta_{n_1-1}^1)$

et plus généralement :

$$\eta_{t+n_i}^i = \sum_{j=1}^{i-1} \sum_{k=1}^{n_i} \alpha_{ijk} \eta_{t+k-1}^j + \sum_{k=1}^{n_i} (\alpha_{iik} \eta_{t+k-1}^i + \beta_{ik}^T u_{t+k-1}) + d_{i-t+n_i}^T u_{t+n_i} \quad (V-100)$$

avec $(\underline{\eta}_0^i)^T = (\eta_0^i, \dots, \eta_{n_i-1}^i)$

... équation dont la configuration est celle d'un problème mono-sortie, là encore grâce à la structure particulière des formes canoniques utilisées, ne présentant des couplages que dans un seul sens. On remarque également que η_t^i n'est une fonction non linéaire que des n_i paramètres α_{iik} .

En conséquence on est ramené pour chaque sortie au problème traité § 2, tant en ce qui concerne la procédure de minimisation de l'erreur de sortie, que la détermination des n_i en fonction de l'évolution de ce critère ou encore en combinaison avec l'erreur d'équation et le placement des pôles et zéros

Sur la base de ce raisonnement, on peut concevoir une série d'algorithmes "ter" obtenue à partir d'algorithmes "bis" en introduisant à l'étape -2- une procédure d'optimisation non linéaire initialisée par les paramètres α_i^* . Cette démarche n'est autre que la généralisation de la procédure d'initialisation automatique développée dans le cas monovariable pour ajuster une séquence de pondération (Cf. § 3.1).

Cependant pour une dimension $n = \sum n_i$ donnée, le critère global d'erreur sur les p sorties ne sera pas minimisé. La seule manière pour y parvenir serait de traiter toutes les sorties simultanément. Une approche prenant en compte les structures canoniques développées Chapitre III consisterait à poser le problème dans les termes suivants :

1) Etant donné la suite $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, N$;

$$\text{minimer } J = \sum_t ||\underline{y}_t - \underline{n}_t|| \text{ avec } n = \sum_{i=1}^p n_i, \text{ donné et } n_i \text{ entiers } \geq 0. \quad (V-101)$$

$\underline{n}_t$ étant calculée composante par composante suivant l'équation (V-100).

2) augmenter n et aller en (1) jusqu'à ce que, par exemple, J minimum soit inférieur à un certain seuil.

Dans le traitement séquentiel des sorties, l'ordre de sélection jouant un rôle, on dispose là d'un moyen pour approcher cet optimum. Il n'existe malheureusement pas de procédure automatique pour effectuer ce classement a priori et il est le plus souvent utopique d'envisager les $p!$ permutations possibles !. Le seul moyen réel dont on dispose est de réaliser soi-même cette sélection en fonction de la signification physique des sorties et des couplages naturels qui en découlent.

4-3 - L'algorithme mixte

Le problème d'optimisation que l'on a évoqué en (V-101) est parfaitement concevable du point de vue numérique, mais il est fort probable que le coût calcul pour le résoudre soit peu en rapport avec l'objectif à atteindre. Par ailleurs, l'approche séquentielle dont on a parlé jusqu'ici possède un caractère, a priori, sous optimal dont il convient de minimiser les effets.

Nous proposons, dans ce sens, une solution "mixte", basée successivement sur un traitement séquentiel des sorties, puis un traitement global simple.

Le principe en est le suivant :

- ne pas remettre en cause la structure du modèle (invariant n_i) obtenu au cours du traitement séquentiel des sorties.
- minimiser globalement l'erreur de sortie par rapport aux paramètres du modèle qui interviennent linéairement dans l'erreur ϵ_t .

Soit donc (Φ, Γ, H, D) le modèle obtenu au cours de la première phase, avec :

$$\begin{cases} \underline{x}_{k+1} = \Phi \underline{x}_k + \Gamma \underline{u}_k \\ \underline{n}_k = H \underline{x}_k + D \underline{u}_k \end{cases} \quad (V-101)$$

où $\underline{n}_k$ est la sortie modèle. Il apparaît clairement que cette dernière est linéaire par rapport à $\underline{x}_0$, Γ et D , par suite il existe une matrice A fonction de Φ , H et $\underline{u}_k$ telle que :

$$\underline{n} = A \times \underline{\beta}$$

avec :

$$\underline{n}^T = \left[\underbrace{\underline{n}_0^T, \dots, \underline{n}_N^T}_{(N+1)p} \right]$$

$$\underline{\beta}^T = \left[\underbrace{\gamma_{11}^T, \dots, \gamma_{1n_1}^T, d_1^T, \dots, \gamma_{p1}^T, \dots, \gamma_{pn_p}^T, d_p^T, x_0}_{m(n+p) + n} \right]$$

Définissons maintenant le vecteur des sorties "objet" constitué de $\underline{y}_t$:

$$\underline{y}^T = [\underline{y}_0^T, \dots, \underline{y}_N^T]$$

La seconde phase de la méthode mixte consiste alors à déterminer les paramètres $\underline{\beta}^*$ solution du problème de moindres carrés :

$$\min_{\underline{\beta}} || A\underline{\beta} - \underline{y} || \triangleq || A\underline{\beta}^* - \underline{y} || \quad (\dot{V}-105)$$

qui n'est autre que le critère projeté $\tilde{J}$ associé au problème (V-101). En d'autres termes la solution que nous venons de développer pourrait être la phase d'initialisation automatique du problème global optimum évoqué en (V-101). Ne pas chercher à aller plus loin est une décision qui doit intervenir en connaissance d'une part, du critère (V-105) et de la dimension du modèle n et d'autre part, du coût de calcul unitaire de $\tilde{J}$ qui, s'il est encore admissible, une fois multiplié par plusieurs centaines, voire quelques milliers (pour le minimiser par rapport à Φ) peut devenir prohibitif. A titre d'exemple indicatif, considérons un modèle 5 entrées, 3 sorties, et 15 états à déterminer à partir d'un échantillon entrées/sorties de 500 points. Le calcul de $\tilde{J}$ consiste alors à résoudre au sens des moindres carrés un système de 1 500 équations à 115 inconnues ! (avec les difficultés que cela peut comporter).

Pour l'instant, nous retiendrons de cette remarque, l'importance qu'il faut accorder à la phase séquentielle au cours de laquelle la structure du système est déterminée et par voie de conséquence l'ordre dans lequel il convient de traiter les sorties.

Nous proposons donc d'opérer de la manière suivante :

- Sélectionner d'abord l'ordre des sorties sur la base de leur signification physique ou des couplages naturels qu'elles peuvent présenter,

ou en s'aidant éventuellement d'une modélisation succincte sortie par sortie.

- Appliquer à ces données l'algorithme mixte que nous allons décrire maintenant.

Algorithme mixte

- 1 - Etant donné la suite $\{\underline{u}_t, \underline{y}_t\}$ $t = 0, \dots, N$; permettant d'identifier un système d'ordre inférieur ou égal à q (donné) ; poser $N' = N - q$.
- 2 - Pour $i = 1, \dots, p$
 - Construire S_q^i (V-89)
 - Triangulariser S_q^i par les transformations de Householder et en déduire $\|\underline{e}^*\|$, $\underline{\alpha}_i^*$ pour un n_i donné d'après (V-93) - (V-94) ou (V-95) - (V-96) suivant que n_i est nul ou positif.
 - Option 1 calculer pour chaque n_i l'erreur de sortie correspondante $\|\underline{\varepsilon}_i^*\|$
 - Option 2 calculer les zéros et pôles pour chaque n_i
 - En déduire la valeur de n_i pour modéliser la $i^{\text{ème}}$ sortie
 - Option 3 minimiser $\|\underline{\varepsilon}_i\|$ par programmation non linéaire avec l'initialisation automatique fournie par le modèle erreur d'équation.
- 3 - Construire la matrice $A(\Phi, H, \underline{u}_k)$, Φ et H étant conformes aux matrices définies dans l'algorithme 5 bis.
 - Résoudre par la méthode de Householder le problème (V-105)

$$\min_{\beta} \|A\beta - \underline{y}\|$$
 - En déduire Γ et D d'après (V-104).

Remarque : Là encore on a raisonné sur la forme canonique (2), il est clair que cette même démarche s'applique à n'importe laquelle des autres formes canoniques observables décrites Chapitre III.

Enfin il est bien évident que l'utilisation de l'entrée impulsion ne change rien au problème ce qui signifie qu'une structure analogue à celle de l'algorithme mixte peut être appliquée aux algorithmes 1 à 4, $\underline{x}_0$ n'intervenant pas puisqu'il est nécessairement nul.

5 - RESULTATS EXPERIMENTAUX (SYSTEMES MULTIVARIABLES)

Nous ne présenterons, ici, que des résultats obtenus sur processus réel, ceci pour deux raisons. La première est qu'en définitive, il s'agit là du véritable contexte dans lequel devront prendre place les algorithmes d'approximation multivariable que nous avons développés. La seconde est qu'il est difficile, en simulation, de prétendre à une analyse expérimentale exhaustive, en raison non seulement du volume de résultats auxquels il faudrait faire référence, mais encore du coût et temps calcul que cela représenterait.

Ceci ne signifie pas pour autant qu'aucune étude en simulation n'ait été faite, ne serait ce que pour vérifier l'exactitude et l'efficacité des modules composant le programme. Cependant plutôt que de présenter des résultats qui seraient, dans ce domaine, inmanquablement incomplets, il nous paraît plus significatif de rapporter ici une des expériences que nous avons menées sur une centrale thermique E.D.F..

Dans le cas présent, il s'agit d'une des tranches de 250 MW (tranche 5) de la Centrale de Nantes-Cheviré. Les résultats d'identification, dont nous allons faire état, prennent place dans le cadre d'un contrat D.G.R.S.T.[†] entre l'E.D.F., la C.E.R.C.I. et le Laboratoire d'Automatique de l'E.N.S.M., contrat ayant pour objectif la commande en ligne de cette tranche, par calculateur numérique.

5.1 - Description de la tranche 5

Cette tranche, d'une puissance installée de 250 MW comprend :

- Un générateur de vapeur
- Un groupe turbo-alternateur
- Un transformateur

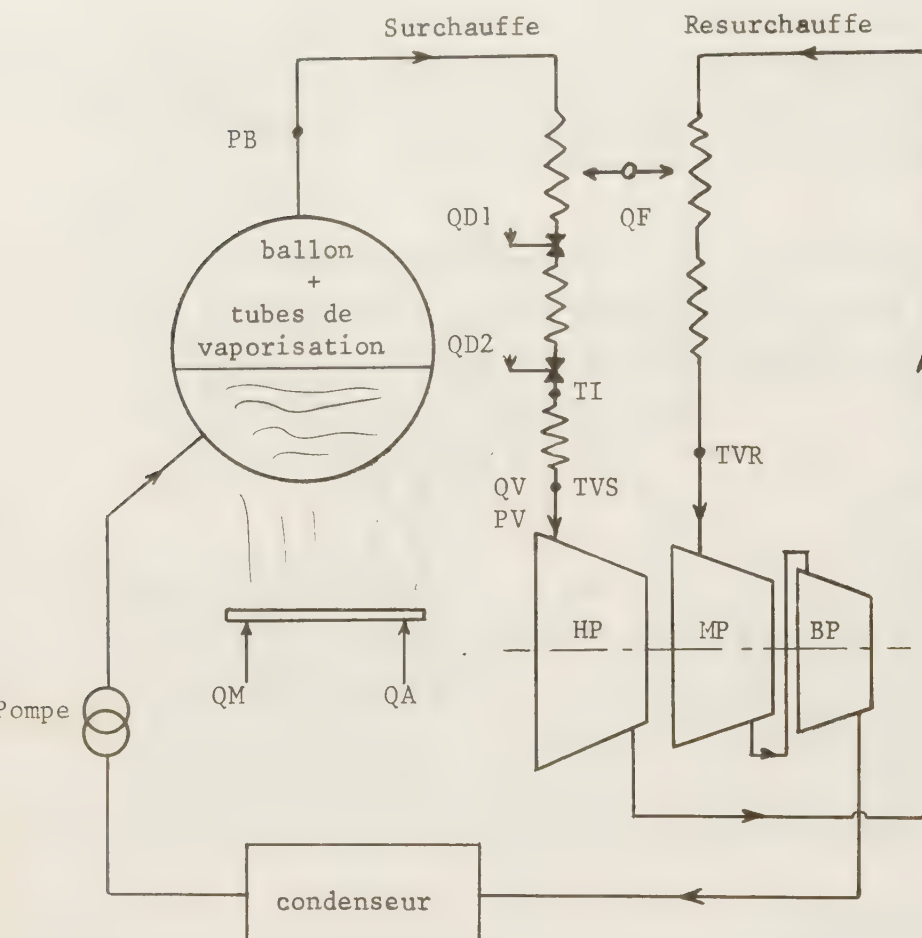
Nous allons nous intéresser plus particulièrement au circuit vapeur dont le générateur a pour principales caractéristiques :

- Débit de vaporisation : 735 T/h
- Température de la vapeur surchauffée : 565°C
- Température de la vapeur resurchauffée : 565°C
- Pression de la vapeur surchauffée : 165 bars

[†] D.G.R.S.T. Action concertée : Comité Automatisation, contrat n° 72.7.0159.

- Pression de la vapeur resurchauffée : 35 bars
- Température de l'eau d'alimentation : 250°C

L'installation répond dans son principe au schéma simplifié suivant (Fig. 15) où l'on a mentionné les points de mesures essentiels.



Notations

PB	: pression ballon
QD1	: débit première injection
QD2	: débit deuxième injection
TI	: température intermédiaire
QV	: débit vapeur
PV	: pression vapeur (H.P.)
TVS	: température de vapeur surchauffée
TVR	: température de vapeur resurchauffée
QF	: répartition fumée
QM	: débit fuel
QA	: débit d'air

Figure 15

5.2 - Définition des grandeurs d'entrées/sorties

5.2.1 - Grandeurs de sortie

Ce sont les variables à contrôler

PV
TVS
TVR

auxquelles on a ajouté une variable intermédiaire mesurée TI qui résulte en réalité de la moyenne de deux mesures TID et TIG relatives à deux surchauffeurs en parallèle. Par ailleurs, au lieu d'utiliser directement PV la pression de sortie, nous avons pris PB, la pression ballon reliée à cette dernière par la relation linéarisée :

$$PV = PB - \alpha QV + \beta$$

où α et β sont des paramètres estimés indépendamment. Cette substitution permet de se dispenser d'identifier une matrice D dans le modèle d'état.

Finalement les sorties retenues sont :

PB ; TI ; TVS ; TVR

l'ordre étant celui où elles interviennent d'une manière naturelle.

5.2.1 - Grandeurs d'entrée

Ce sont les commandes :

CQM

CQD

CQF

Les signaux d'entrée considérés ne sont donc pas directement les débits fuel (QM) d'injection (QD1 + QD2) et de répartition fumée (QF). En effet le débit fuel observé résulte d'un asservissement local de début fuel et c'est la consigne de cet asservissement, soit CQM, que nous considérons comme grandeur d'entrée. De même la commande de répartition des fumées est la consigne CQF de l'asservissement correspondant. En ce qui concerne les deux injections, elles sont commandées à partir d'une même consigne d'injection CQD par l'intermédiaire d'un dispositif complexe dans lequel intervient des limitations de sécurité. Sans entrer dans les détails, nous indiquerons simplement que CQD est la somme de la consigne sur la deuxième injection et du débit de la première (CQD = CQD2 + QD1).

A ces trois entrées "naturelles" nous avons ajouté une quatrième grandeur, le débit vapeur QV. Elle représente la perturbation principale de la tranche, et est en fait, ici, l'image de la puissance demandée par le réseau. En conséquence, ce qu'on aurait pu considérer à première vue comme une sortie, se trouve être une variable indépendante définissant en réalité une entrée, non manipulable.

En définitive, il s'agit donc de modéliser le système 4×4

suivant :

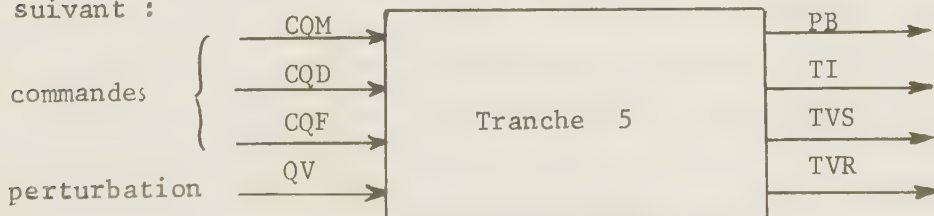


Figure 16

Remarque : Nous n'avons pas fait intervenir le débit d'air QA, celui-ci étant pratiquement proportionnel au débit fuel QM.

5.2.3 - Signaux d'excitation

Pour des raisons de sécurité évidentes, les services techniques de la centrale préfèrent que la tranche soit maintenue en boucle fermée sous le contrôle des régulations existantes. L'excitation des grandeurs d'entrées à partir desquelles a été effectuée la modélisation, a donc été obtenue indirectement en agissant sur les consignes des asservissements. En fait il a été possible d'ouvrir certaines boucles sauf celle de la pression qui est instable en boucle ouverte comme le montreront d'ailleurs certains résultats de l'identification.

L'acquisition des données d'entrées/sorties correspondantes, a été faite avec un pas de 5 secondes. Les 10 voies de mesure concernées sont, rappelons le, CQM, QV, CQD2, QD1, CQF pour les entrées et PB, TID, TIG, TVS et TVR pour les sorties. Ces quantités ont été exprimées en volts sur la base du tableau de correspondance suivant :

Grandeurs	0	10 volts
TI, TVS, TVR	300°C	600°C
CQD	0	60 T/h
CQM	0	70 T/h
CQF	0	100 %
QV	0	800 T/h
PV, PB	0	200 bars

Tableau 13

Précisons enfin que le pas d'échantillonnage utilisé pour la modélisation effective du système a été ramené à 50 s en remplaçant chaque groupe de 10 mesures par sa valeur moyenne.

Nous présentons maintenant les modèles que l'on a obtenus à l'issue de deux essais effectués respectivement le 26 février et le 30 septembre 1974.

5.3 - Identification de la tranche 5

5.3.1 - Essai du 26 février 1974

La tranche fonctionnait alors en télé réglage autour d'un régime nominal de 210 MW. Les excitations ont été obtenues en agissant sur les consignes d'asservissement de TVS, TVR et PV.

Les caractéristiques des données sont les suivantes :

- durée de l'essai : 3 h 55 mn 55 sec
- période d'échantillonnage : 5 sec
- nombre de mesures par voies : 2 832
- nombre de points utilisés (avec un pas de 50 sec) : 283

Le tracé relatif aux enregistrements des 10 voies utilisées pour identifier le système, a été reproduit à la Fig. 17 pour les entrées et Fig. 18 pour les sorties.

L'application de l'"algorithme mixte" a conduit aux résultats suivants , en notant $||\hat{\underline{\epsilon}}||$ l'erreur de sortie minimum, n_i le nombre d'états introduit pour chaque sortie, p le nombre de paramètres relatif à chaque sous système mono-sortie, p_t le nombre total de paramètres intervenant dans le modèle, $||\hat{\underline{\epsilon}}||_g$ l'erreur globale sur les sorties, CPU le temps calcul.

	$ \hat{\underline{\epsilon}} $	n_i	p	p_t	$ \hat{\underline{\epsilon}} _g$	CPU
PB	45 %	3	15	70	37 %	4,4583 heures
TI	29 %	3	18			
TVS	29 %	1	11			
TVR	49 %	2	17			

Tableau 14

Ces résultats sont complétés au Tableau 15 par les valeurs des pôles du modèle et par le tracé d'une part, des 16 réponses indicielles caractérisant le système (Fig. 19) et d'autre part de celui des 4 sorties, modèle et objet (Fig. 20).

	pôles du modèle
PB	1.018 ; 0.5266 ; -0.1278
TI	0.7661 ; 0.2243 ± i 0.3524
TVS	0.5848
TVR	0.8143 ± i 0.0856

Tableau 15

Avant d'apporter quelques commentaires à ces résultats, nous allons d'abord présenter ceux obtenus au cours d'un second essai effectué récemment.

5.3.2 - Essai du 30 septembre 1974

Celui-ci se caractérise par un fonctionnement en asservissement de puissance autour de 245 MW, à l'exception d'un créneau de 20 mn à 235 MW. Les autres excitations ont été réalisées à partir de variation des points de consigne sur CQD, CQF et PVC (consigne de PV). Signalons encore que les températures TVS et TVR étaient alors en boucle ouverte.

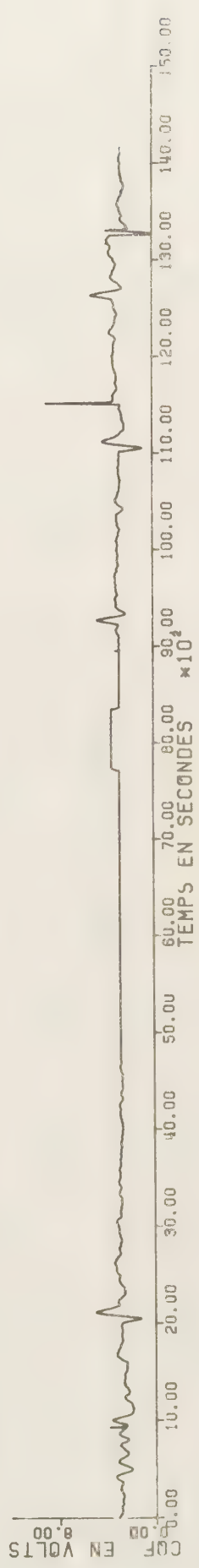
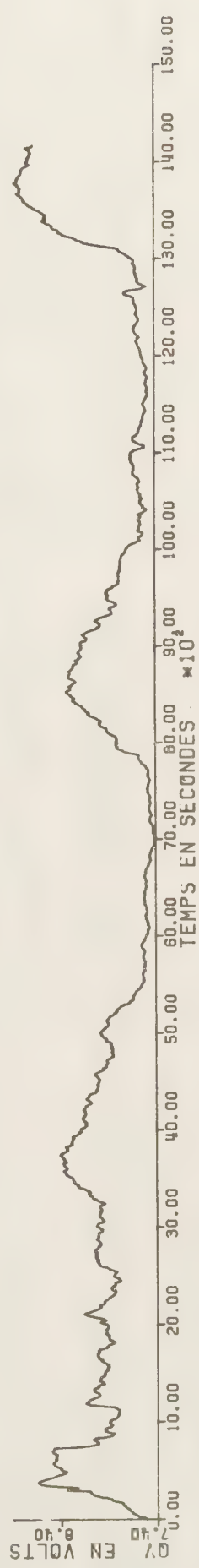
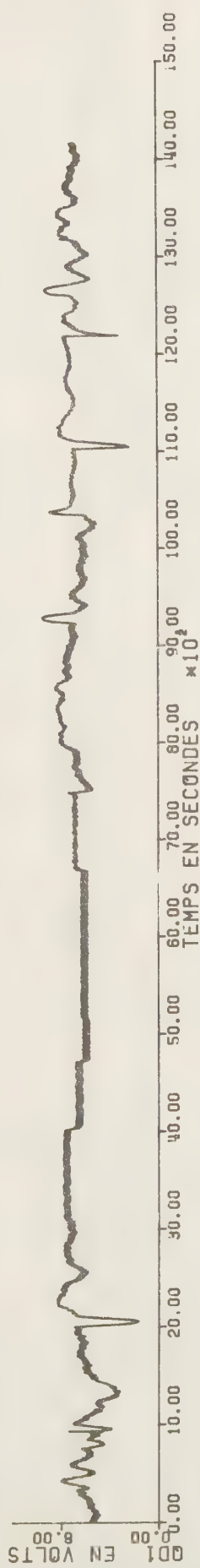
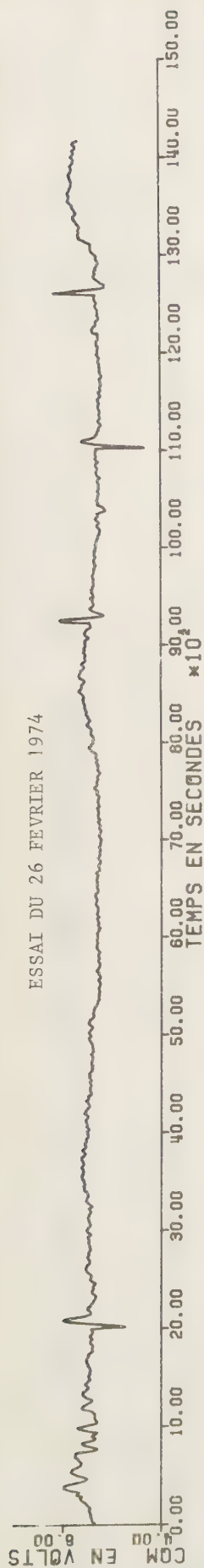
Les autres caractéristiques de l'essai sont :

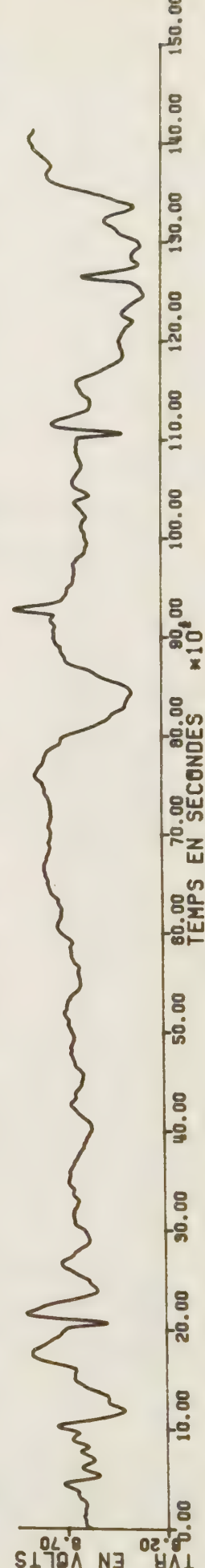
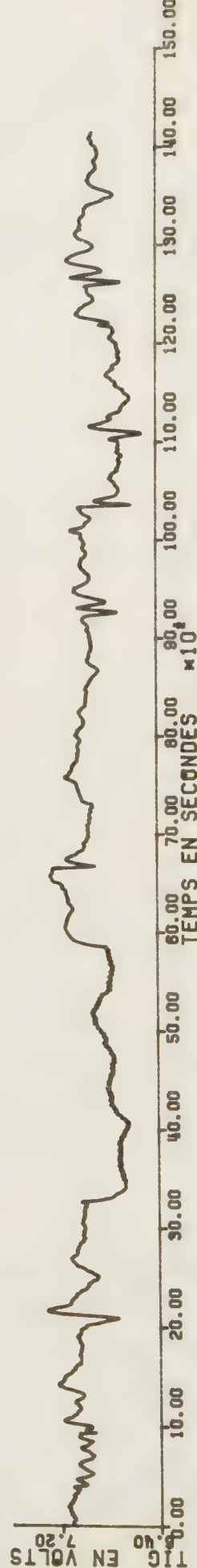
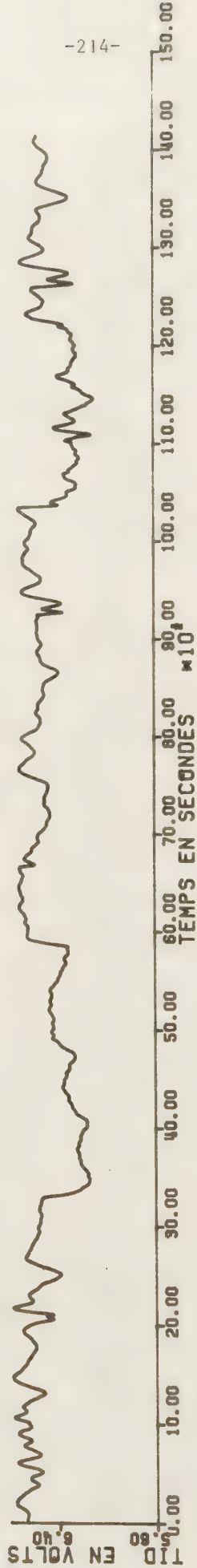
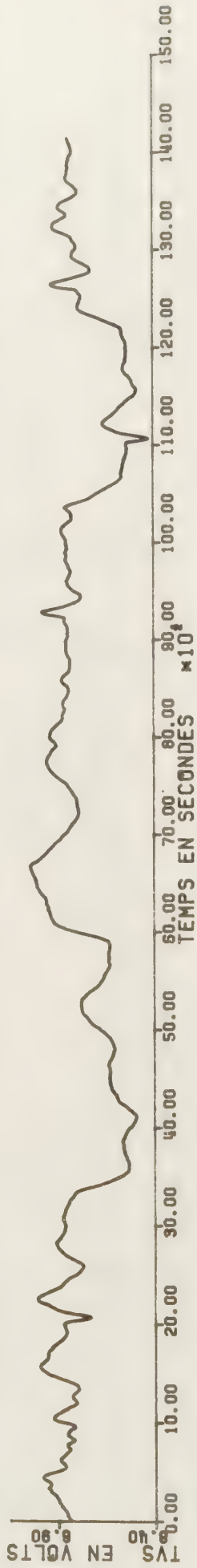
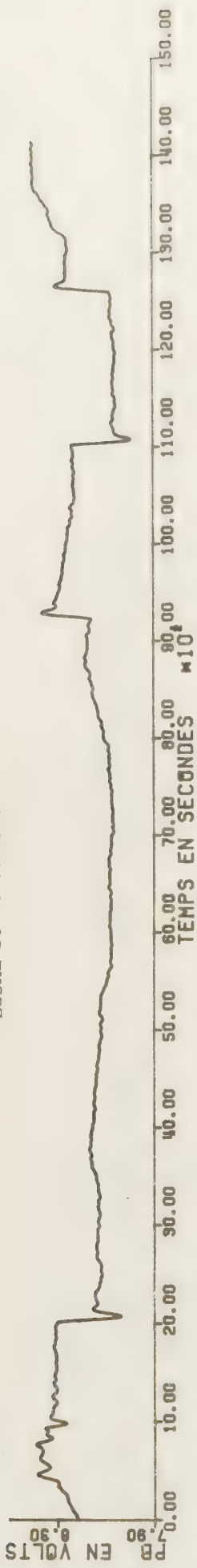
- durée de l'essai : 3 h 35 mn 25 sec
- période d'échantillonnage : 5 sec
- nombre de mesures par voies : 2586
- nombre de points utilisés (avec un pas de 50 sec) : 258

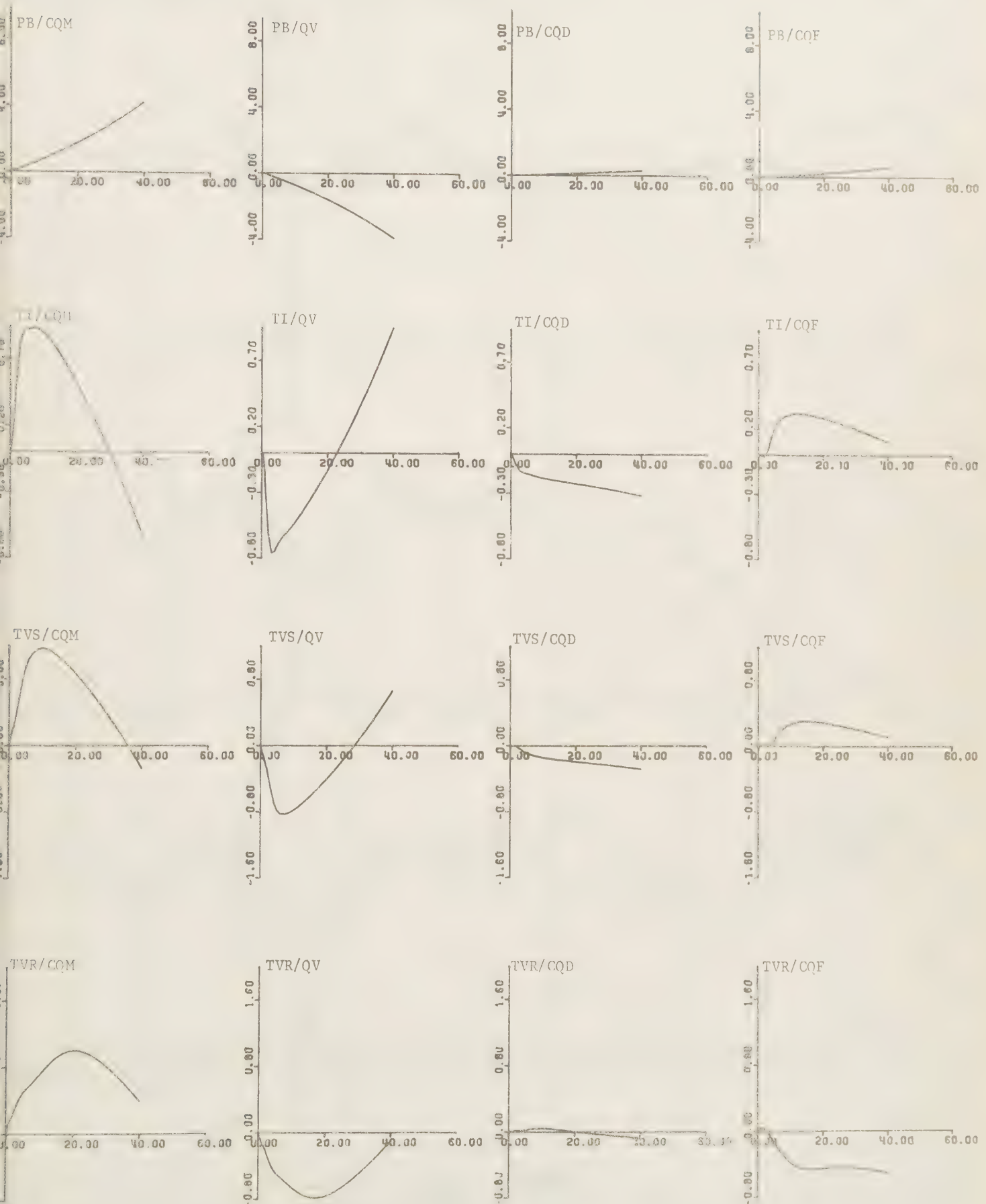
Comme précédemment, on a reproduit les enregistrements des 10 voies de mesures à partir desquelles ont été constituées les entrées du modèle (Fig. 21) et ses sorties (Fig. 22).

Les résultats chiffrés, concernant ce second modèle, sont reportés au Tableau 16 et 17, le premier concernant les caractéristiques générales du modèle, le second étant réservé aux valeurs obtenues pour les pôles du système.

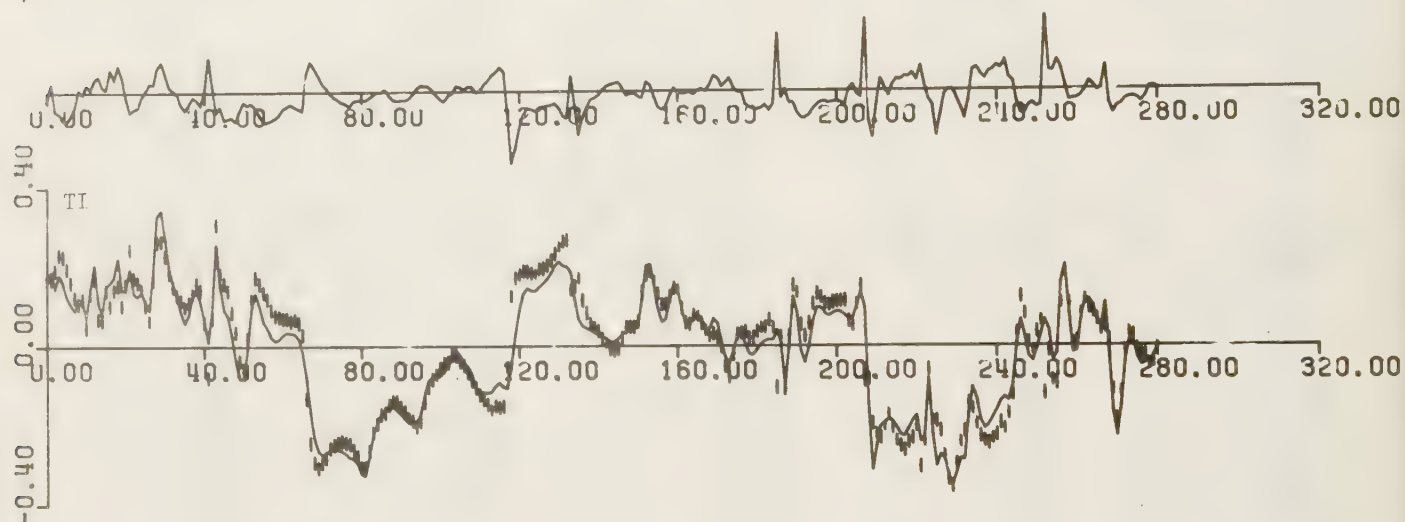
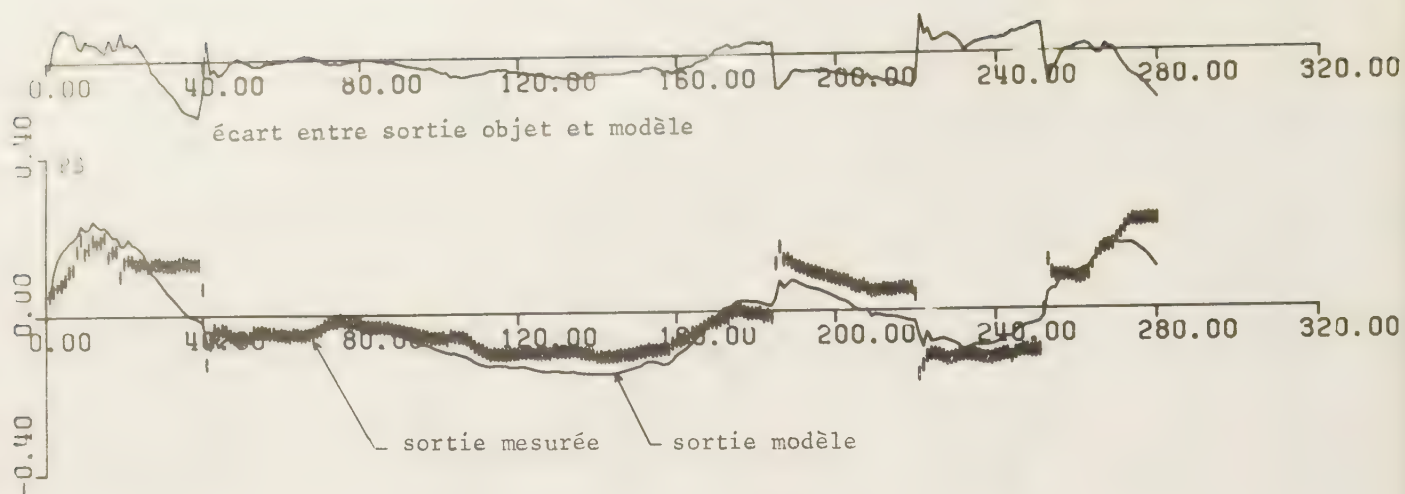
ESSAI DU 26 FEVRIER 1974

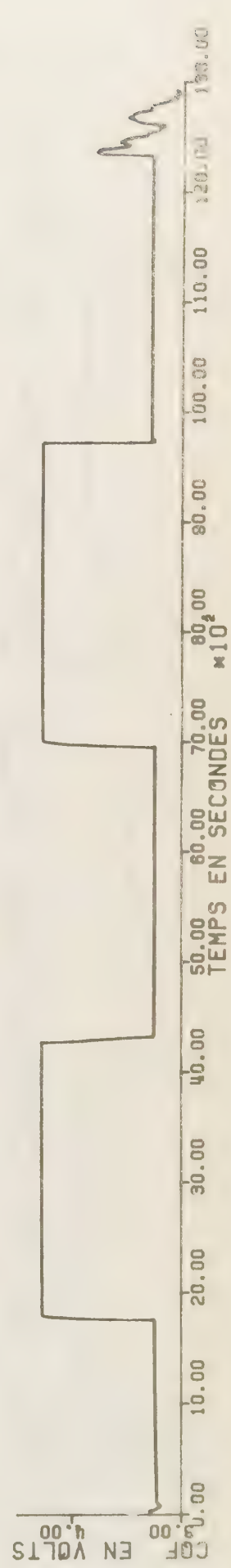
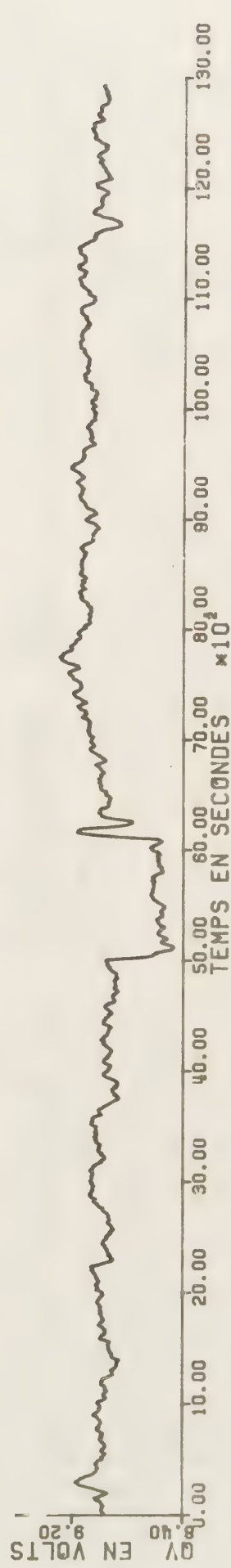
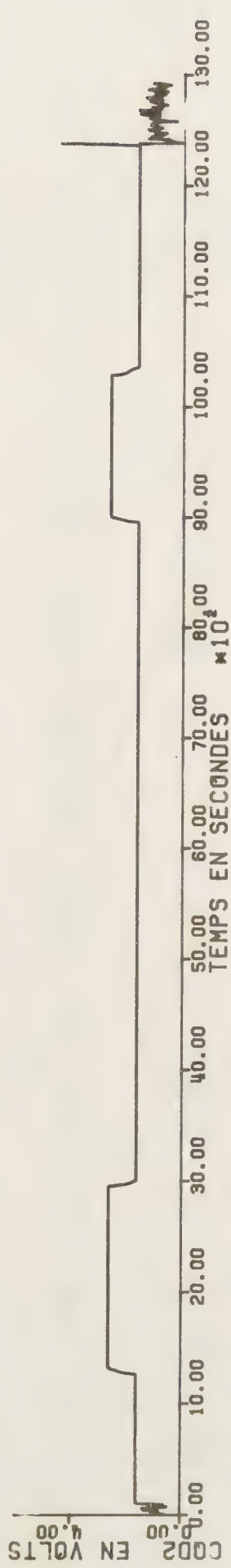
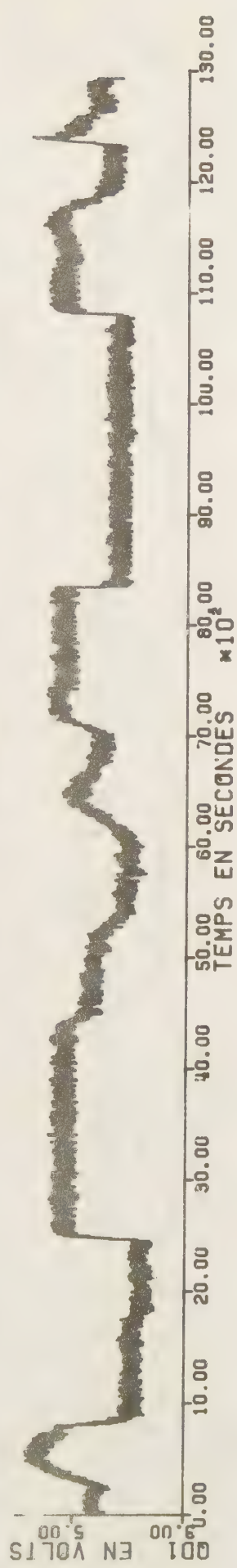
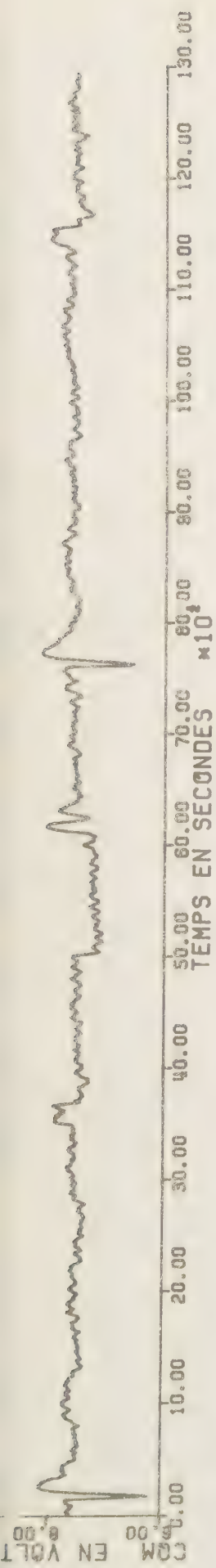


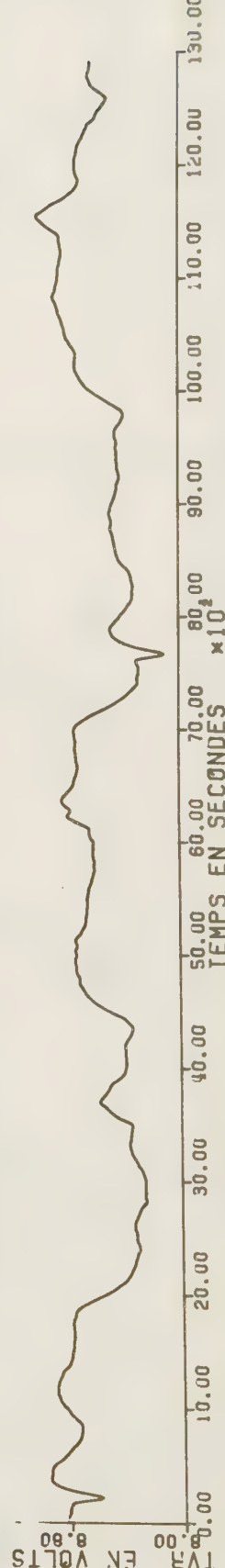
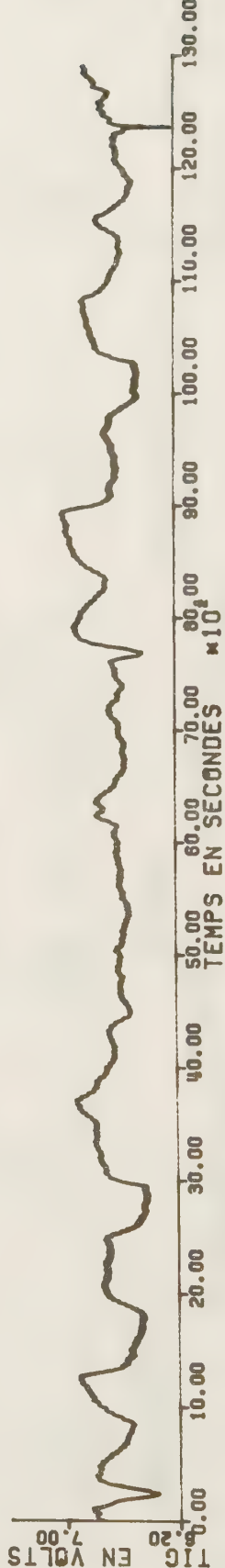
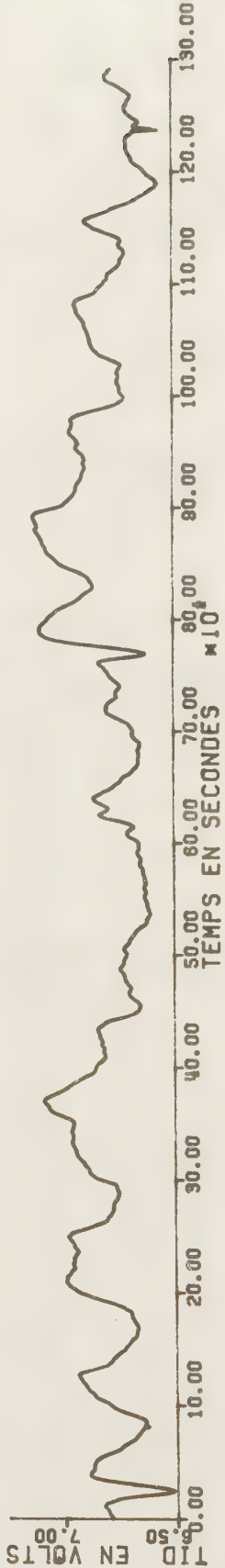
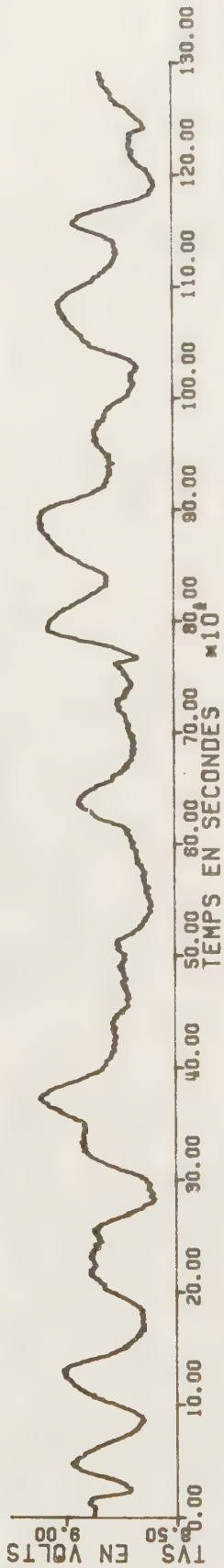
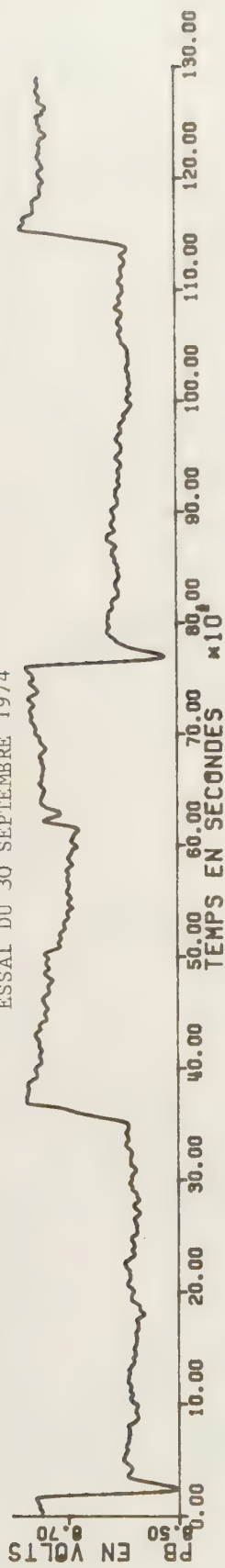


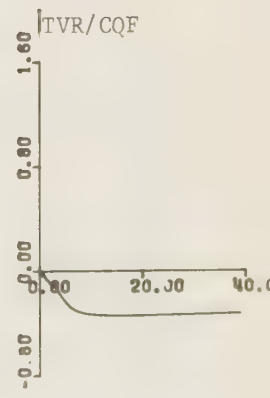
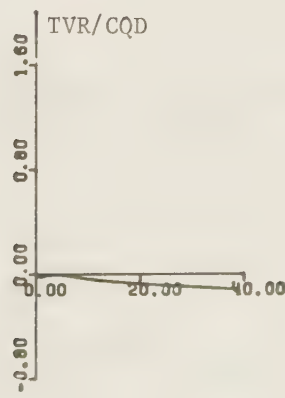
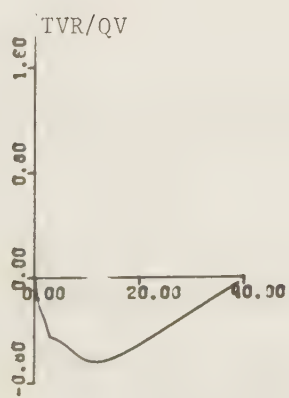
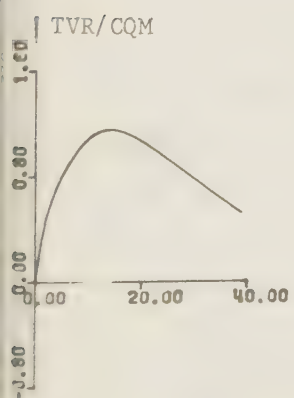
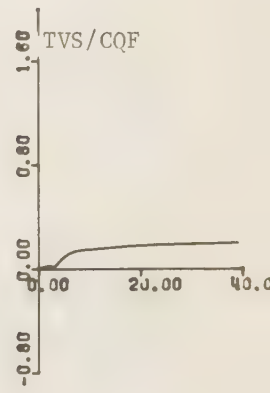
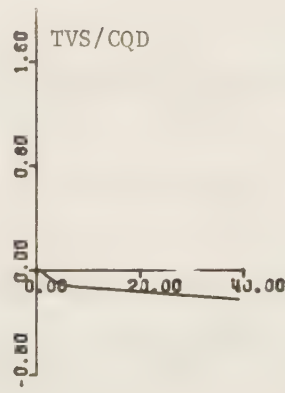
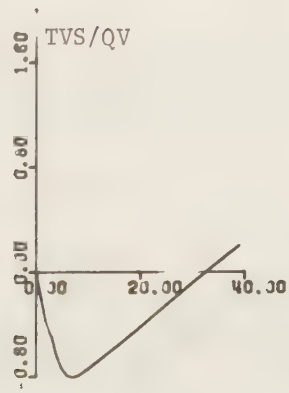
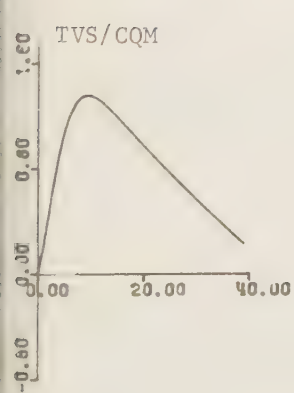
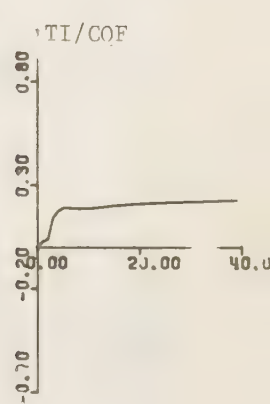
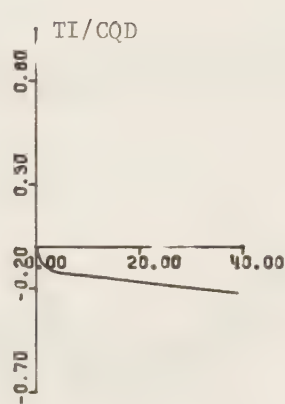
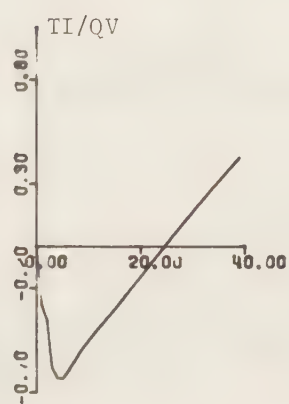
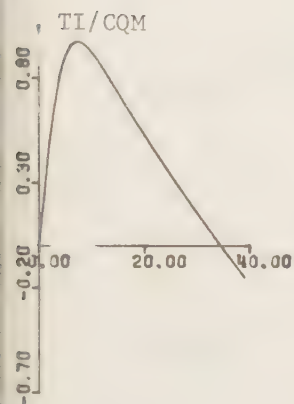
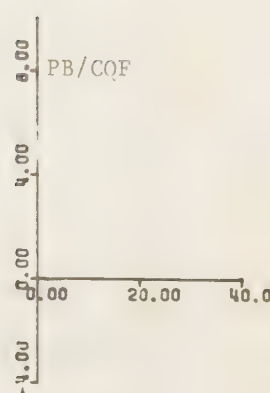
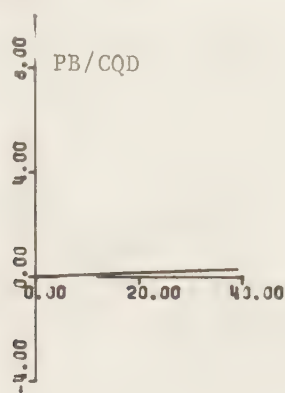
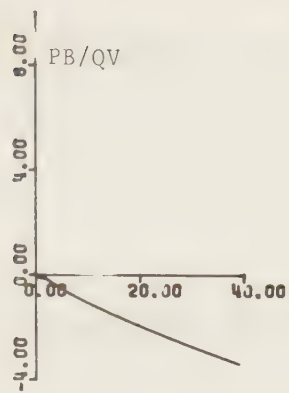
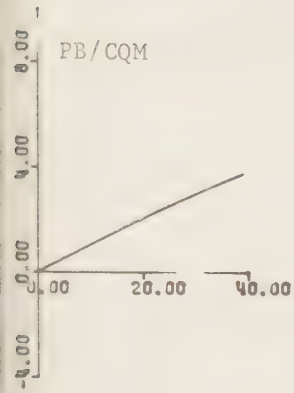


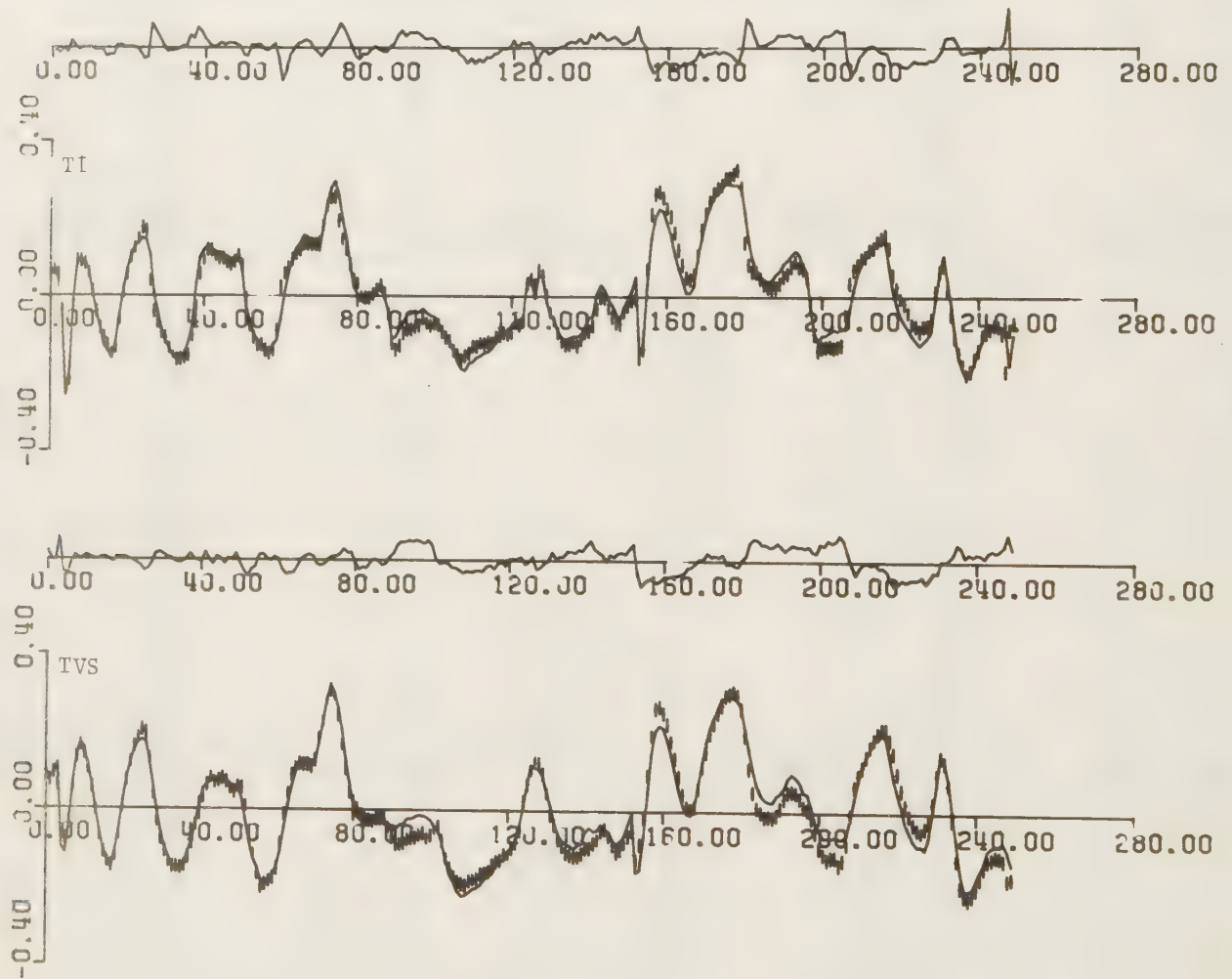
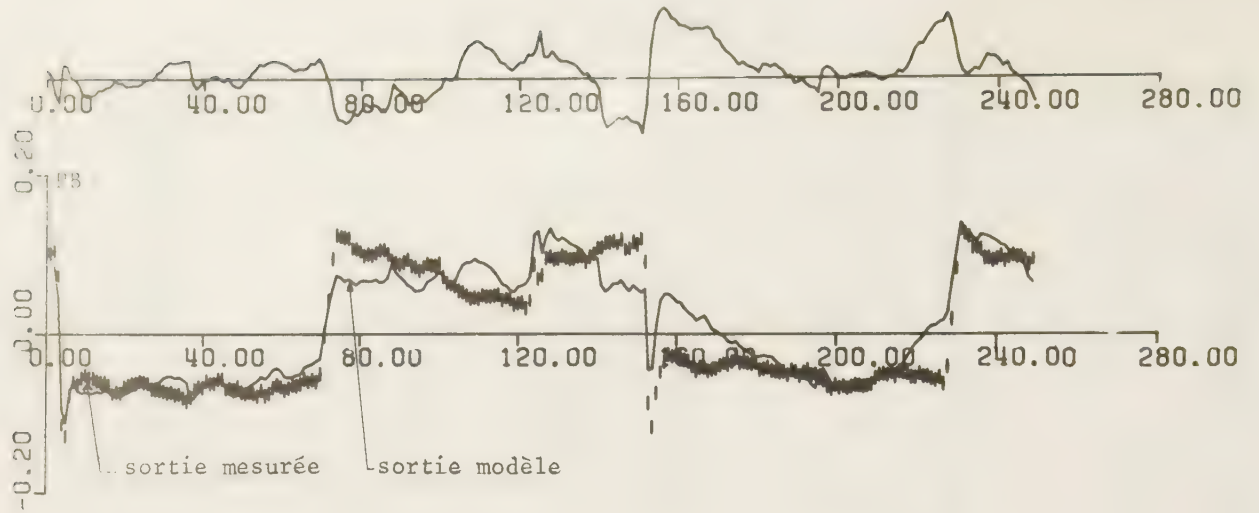
ESSAI DU 26 FEVRIER 1974











	$ \hat{\underline{\epsilon}} $	n_i	p	p_t	$ \hat{\underline{\epsilon}} _g$	CPU
PB	42 %	4	20	65	21 %	3,5437
TI	23 %	2	14			
TVS	22 %	1	11			
TVR	15 %	1	12			

Tableau 16

	pôles du modèle
PB	0.9909 ; 0.7321 ; 0.3904 ; -0.3320
TI	0.6017 ; 0.5018
TVS	0.5015
TVR	0.7789

Tableau 17

Là encore on a complété ces résultats par le tracé des réponses indicielles (Fig. 23) et par celui des sorties objet et modèle (Fig. 24).

5.3.3 - Commentaires sur les résultats expérimentaux

Les pourcentages d'écart entre les sorties modèle et objet, relativement importants dans certains cas, ne mettent pas en cause la manière dont ont été traitées les données mais simplement les conditions d'expérience qui sont telles que certaines variations sur les sorties ne peuvent être imputées aux entrées. Il existe par exemple des dérives (variation du régime permanent due par exemple à un changement de débit d'air) qui n'ont pas été modélisés et qui se traduisent, notamment pour la pression, par des écarts importants entre objet et modèle. A ceci il faut ajouter le faible niveau de variation des sorties par rapport au niveau de bruit et de perturbations. En effet en fonctionnement normal, les variations de températures ou de pression sont de l'ordre de 2° ou 3° et 1 bar respectivement pour n'atteindre que 6° à 8° et 3 à 4 bars au cours des essais (excitation des entrées par variation des points de consignes des asservissements).

En ce qui concerne la stabilité de la chaîne de pression, le fait de trouver pour un modèle un pôle légèrement supérieur à 1(1.018) et dans l'autre cas à la limite de la stabilité (0.991), est relativement peu important dans la mesure où les réponses indicielles coïncident approximativement sur un laps de temps suffisant, ce qui est le cas ici (cf Figures 19 et 23). En effet la tranche étant destinée à fonctionner en boucle fermée, le contrôle se fera alors de la même manière dans l'un et l'autre cas.

Plus généralement, on pourra constater (Figures 19 et 23) que les 16 réponses indicielles ont même allure, et sont presque superposables, en notant toutefois des réponses plus rapides pour l'essai du 30 septembre, réalisé à plus haute charge, ce qui s'explique par l'accélération des processus d'échange thermique dû à un débit vapeur accru.

A propos de la structure des modèles, elle s'est imposée par la méthode couplage dans un seul sens. Signalons cependant que des résultats comparables ont été obtenus en opérant bloc par bloc sur la base d'une étude physique du système. Notre objectif n'étant pas ici une étude comparative des différentes approches que l'on pouvait donner à ce problème d'identification, nous ne nous étendrons pas sur ce point.

Il nous paraît plus important de faire remarquer la similitude des deux modèles (Cf. Figures 19 et 23) déterminés dans des conditions totalement différentes et à plus de 6 mois d'intervalle. Bien qu'il eut été possible de traiter simultanément ces deux séries de mesures (essais du 26 février plus du 30 septembre), leur traitement distinct et les résultats obtenus constituent une validation du modèle, confirmée indépendamment par d'autres expériences : identification de bloc élémentaire par filtrage non linéaire.

6 - COMMENTAIRES SUR LES METHODES NUMERIQUES -

La mise en oeuvre sur ordinateur numérique est le plus souvent le prolongement naturel des travaux d'un chercheur. Néanmoins la réalisation de programmes de calculs concrétisant certains résultats théoriques demande parfois un travail tel qu'il pourrait constituer, à lui seul, un secteur d'activité distinct.

Dans le cadre de ce mémoire, consacré aux problèmes de réalisation et d'approximation des systèmes linéaires multivariables, la mise en place des programmes mettant en oeuvre l'ensemble des algorithmes que nous avons décrits, représenterait un travail d'environ 4 années ingénieurs. On conçoit donc facilement là encore, qu'il n'était guère envisageable d'illustrer tous les points que nous avons développés.

La quantité de travail à laquelle nous venons de faire allusion, peut paraître importante dans la mesure où la plupart des problèmes numériques à résoudre semblent plus ou moins classiques. Ils le sont effectivement d'un point de vue conceptuel, mais leur dimension ou/et la nature des données les rendent néanmoins difficiles à appréhender par des techniques courantes et nécessitent en pratique l'emploi de méthodes numériques très élaborées. Rappelons simplement pour mémoire l'extraordinaire sensibilité du critère erreur de sortie intervenant dans la modélisation des séquences de pondération ou la dimension de certains systèmes linéaires ($m > 1000, n > 100$) à résoudre au sens des moindres carrés.

Pour mieux comprendre les développements que nous avons donnés à ce domaine, il est nécessaire de resituer le problème dans le contexte où il s'est posé.

La première version du programme d'approximation multivariable à partir d'entrées/sorties quelconques a été écrite au cours de l'année 1972. Le laboratoire avait alors accès à un I.B.M. 360-30 64 K. En raison de l'exiguïté de sa mémoire centrale, il était hors de question d'effectuer les calculs en double précision, ce qui ramenait à 6 décimales la précision machine. Compte tenu de la quantité de calculs qui pouvait intervenir dans un tel programme, il était indispensable de minimiser, tant que faire se peut, la propagation d'erreur d'arrondi.

Pour y parvenir, on a conçu l'ensemble du programme sous forme modulaire. De cette conception même, il ressort que l'opération de base est le produit scalaire dont l'imprécision (en valeur relative) est bien connue. A titre d'exemple, imaginons un produit scalaire qui aurait conduit à effectuer la somme des nombres suivants :

$$\begin{aligned} A &= 0.100.10^0 & ; B &= 0.400.10^{-3} & ; C &= -0.500.10^{-1} \\ D &= -0.700.10^{-3} & ; E &= -0.500.10^{-1} & ; F &= 0.700.10^{-3} \end{aligned}$$

... au moyen d'un calculateur n'ayant que 3 décimales flottantes.

On vérifie alors facilement que :

$$S_1 = A + B + C + D + E + F = 0.0 \text{ (exactement en machine)}$$

Par ailleurs, si ces mêmes nombres s'étaient présentés dans un ordre différent, on aurait obtenu par exemple :

$$S_2 = B + F + A + C + D + E = 0.300.10^{-3}$$

$$S_3 = A + B + F + C + E + D = -0.700.10^{-3}$$

... alors qu'en réalité, le vrai résultat est $S = 0.400.10^{-3}$! Il est évident qu'une telle disparité dans les résultats peut-être quelque fois catastrophique. Une première solution (classique) pour remédier à ce genre de problème consiste à accumuler les sommes partielles des flottants simple précision A,B,..., dans un mot double précision. Dans le cadre de l'exemple ci-dessus on trouverait effectivement le bon résultat quel que soit l'ordre dans lequel la somme des nombres A,B,... serait réalisée. Cependant, cette procédure ne fait que diminuer les risques précédents ne permettant pas de garantir une quelconque précision relative du résultat. Dans le cas présent où la précision standard n'est que de 6 décimales, on a préféré adopter une technique, plus lourde il est vrai, mais garantissant au résultat une précision relative identique à celle de la machine ! Cette solution a été développée sur la base d'un article de MALCOLM [42] ; son principe consiste à ventiler les nombres, dont on a à faire la somme algébrique, non pas dans un accumulateur double précision, mais dans un *ensemble* de tels accumulateurs, le dispatching s'effectuant suivant leur ordre de grandeur (contenu de la caractéristique). Cette opération étant effectuée, le produit scalaire est ensuite obtenu en faisant la somme du contenu *des* accumulateurs.

Un deuxième module, d'un niveau supérieur, concerne la triangularisation d'une matrice par la méthode de Householder, opération qui joue un rôle fondamental, ainsi qu'on a pu s'en rendre compte au cours des Chapitres IV et V. Sa programmation s'inspire dans son principe de la version algol décrite dans [10]. Toutefois, en raison de la dimension des matrices à traiter et de la taille mémoire dont on disposait, la mise en oeuvre se fait colonne par colonne, le reste de la matrice étant sur fichier disque. La méthode de Householder se prête bien à ce type de traitement séquentiel, de sorte que cette technique n'apporte aucune difficulté supplémentaire par rapport à l'approche classique [10]. Par ailleurs, l'usage du module, produit scalaire, en simplifie l'écriture et la compréhension.

D'un niveau parallèle, on a défini une autre fonction, fréquemment utilisée et intervenant dans la détermination de paramètres du système, il s'agit de la résolution d'un système triangulaire.

Ce module ne présente aucune difficulté sur le plan numérique et se réduit à sa plus simple expression par l'emploi de l'opérateur produit scalaire. Notons ici que dans le cadre de la réalisation exacte d'un processus, le système à résoudre est généralement mal conditionné puisque par définition, ses colonnes sont à la limite de la dépendance linéaire. C'est pourquoi, on a introduit une procédure de raffinement itératif [10] au cours de laquelle on teste si le

conditionnement du système est compatible avec la précision machine.

Ce sont là les principaux éléments constitutifs de la version de base du programme d'approximation multivariable dont les très nombreux tests ont permis de vérifier expérimentalement l'excellente stabilité numérique de l'ensemble.

Avant de poursuivre, nous voudrions faire quelques remarques concernant une autre approche possible des problèmes de mémoire. Celui-ci avait été résolu en utilisant explicitement un fichier disque. Il est clair que maintenant la question se pose de moins en moins au niveau de la conception d'un programme depuis l'exploitation des systèmes à mémoires virtuelles. Par ailleurs, tout ce qui a été développé Chapitres IV et V concernant la détermination de n_1 et de l'évolution de l'erreur d'équation, pourrait être transposé de la matrice concernée (notée ici S) à la matrice carrée $S^T S$ calculée en double précision. Dans ces conditions, la triangulation par la méthode de Householder de $S^T S$ ne poserait plus de problème de taille mémoire, la condition pratique étant que $S^T S$ tienne en mémoire centrale (en double précision). Enfin dans la phase de minimisation de l'erreur globale de sortie où interviennent de grands systèmes à résoudre au sens des moindres carrés, on pourrait avoir recours à une triangularisation, itérative par rapport au nombre d'équations, qui est à rapprocher de la méthode classique de moindre carré itératif dont la formulation est bien connue et largement utilisée. La solution optimale d'un point de vue temps calcul consiste à traiter simultanément le nombre maximum d'équations, compatible avec le volume mémoire dont on dispose. Dans un ouvrage récent de LAWSON et HANSON [43] on indique à ce sujet que le temps de calcul varie dans un rapport 1 à 2 entre le traitement global et le traitement équation par équation.

Ce sont là des suggestions qui n'ont pas encore été exploitées mais qui pourront faire l'objet d'une nouvelle version du programme.

Le "package" utilisé pour effectuer les essais présentés au § précédent est issu de la version dont on a parlé un peu plus haut, à laquelle on a ajouté d'autres modules permettant en particulier de calculer les zéros et les pôles, de résoudre le problème de minimisation de l'erreur de sortie par programmation non linéaire, d'obtenir les séquences de pondération et réponses indicielles du modèle ainsi que les gains statiques. Signalons également que le programme a été conçu de telle sorte qu'on puisse traiter simultanément plusieurs "batches"

de mesures. L'ensemble représente un peu plus de 2 000 instructions et occupe 200 K ou 100 K suivant que l'on fait usage ou non du recouvrement possible de certains modules ("overlay") et dans des conditions permettant de traiter des systèmes ayant jusqu'à 10 entrées, 10 sorties et 35 états avec la possibilité d'estimer 130 paramètres simultanément.

Exploité, modifié, amélioré depuis plusieurs années, ce programme représente un outil puissant dont les performances sont pour l'instant essentiellement limitées par le temps de calcul lié au système informatique dont on dispose.

7 - CONCLUSION

La modélisation d'un système physique, quel qu'il soit, s'exprime nécessairement en termes d'approximation. Il s'agit d'un problème fondamental, puisque c'est celui que l'on est amené à traiter en pratique.

Modèle et objet ne peuvent être confondus, par définition, il s'avère nécessaire de préciser d'abord dans quels termes l'un est une bonne approximation de l'autre. On ne peut malheureusement pas donner de réponse générale à ce problème préliminaire, ce qui explique la multiplicité des critères s'appliquant eux-mêmes à des objectifs distincts (modèle de connaissance, modèle de conduite, par exemple).

En se limitant d'abord aux systèmes monovariables (ou monosortie), nous avons pu constater les difficultés que pouvait présenter le choix d'un critère, mais encore le calcul d'un modèle optimal par rapport à certains d'entre eux. Rappelons l'extrême sensibilité de l'erreur de sortie par rapport aux pôles du modèle.

Lorsqu'on a abordé ensuite les systèmes multivariables, la complexité du problème a pris une autre dimension d'une part, sur le plan numérique, en raison de l'augmentation rapide du nombre de paramètres, et d'autre part, au niveau de la détermination d'une structure "optimale" du modèle.

Nous avons alors évoqué ce qui pourrait être une solution de référence, mais il est apparu rapidement qu'en pratique, on ne pouvait prétendre à une telle approche.

Le compromis que nous avons retenu, et qui faisait l'objet de l'"algorithme mixte", consistait à ne pas remettre en cause la structure déterminée au cours d'un traitement séquentiel des sorties, puis de compenser le caractère approché de cette solution par une seconde phase globale appropriée.

Les excellents résultats obtenus, au cours de plusieurs campagnes d'essais sur centrale thermique sont là pour témoigner de l'intérêt et de l'efficacité d'une telle approche, face à des processus industriels.

B I B L I O G R A P H I E

- [1] R.G. WILSON
D.G. FISHER
D.E. SEBORG "Model reduction for discrete-time dynamic System"
Int. J. Control, 1972, Vol. 16, n° 3, pp. 549-558.
- [2] Y.P. SHIH
W.T. WU "Simplification of Z transfer function by continued fractions"
Int. J. Control, 1973, Vol. 17, n° 5, pp. 1089-1094
- [3] C.F. CHEN
L.S. SHIEH "A novel approach to linear model simplification"
Int. J. Control, 1968, pp. 561-570.
- [4] Y. SHAMASH "Continued fraction methods for the reduction of discrete-time dynamic systems"
Int. J. Control, 1974, Vol. 20, n° 2, pp. 267-275.
- [5] M.J. BOSLEY
F.P. LEES "A survey of simple transfer function derivation from high order state variable models".
Automatica. Vol. 8, pp. 765-775, 1972.
- [6] D. MITRA "On the reduction of complexity of linear dynamical models-results of computer studies"
U.K.A.E.A. Report R 535, 1967, U.K.
- [7] K.J. ASTROM "Lectures on the identification problem the least squares method"
Report 6806, sept. 1968. Lund Institute of Technology.
- [8] N.K. SINHA
W. PILLE "A new method for reduction of dynamic systems"
Int. J. Control, 1971, Vol. 14, n° 1, pp. 111-118.
- [9] N.K. SINHA
G.T. BEREZNAI "Optimum approximation of high order systems by low-order models".
Int. J. Control, 1971, Vol. 14, n° 5, pp. 951-959.
- [10] P. BUSINGER
G.H. GOLUB "Linear least squares solutions by Householder transformations"
Num. Math. 7, p. 269-275 (1965)
- [11] R. HOOKE
T.A. JEEVES "Direct search solution of numerical and statistical problem".
J.A.C.M. April 1961, Vol. 8, n° 2, pp. 212-229.
- [12] H.H. ROSENBROCK "An automatic method for finding the greatest or least value of a function".
Comp. J. 3(1960), pp. 175.
- [13] J.A. NELDER
R. MEAD "A simplex method for function minimisation"
Comp. J. 7 §1965) p. 308.
- [14] P.E. GILL
W. MURRAY "Quasi Newton Methods for unconstrained Optimization"
J. Inst. Math. Applie. (1972) 9, pp. 91-108.

- [15] R. FLETCHER "FORTRAN Subroutines for minimization by quasi-Newton methods"
A.E.R.E. R 7125, (1972) U.K.A.F.A. England
- [16] D.W. MARQUARDT "An algorithm for least squares estimation of non linear parameter".
J. Soc. Indust. App. Math. 11 (1963), p. 431.
- [17] R. FLETCHER "A modified Marquardt subroutine for non linear least square".
A.E.R.E. R 6799 (1971) U.K.A.E.A. England
- [18] G.H. GOLUB "The differentiation of pseudo inverses and non linear least squares problems whose variables separate"
SIAM J. Num. Anal. Vol. 10, n° 2, april 1973.
- [19] A. ALBERT "Regression and the Moore Penrose pseudoinverse"
Academic Press, 1972, New-York. London
- [20] F.T. KROGH "Efficient implementation of a variable projection algorithm for non linear least squares problems.
C.A.C.M. March 1974, Vol. 17, n° 3, pp. 167-169.
- [21] K.J. ASTROM "A numerical method for the evaluation of complex integrals".
E.I. JURY
R.G. AGNIEL
I.E.E.E. Tr Aut. Cont. August 1970, pp. 468-471.
- [22] K. MASDEN "A root finding algorithm based on Newton's method"
Bit 13 (1973) pp. 74-75.
- [23] J.F. LAFAY "Séparation des variables dans la minimisation de l'erreur de sortie. Application à l'identification à partir de séquence de pondération".
D.E.A. Janvier 1975. E.N.S.M. Nantes
- [24] C.M. WOODSIDE "Estimation of the order of linear systems"
Automatica Vol. 7, pp. 727-733, 1971.
- [25] H. UNBEHAUEN "Application of different statistical tests for the determination of the most accurate order of the model in parameter estimation".
G. GOHRING
Congrès IFAC La Haye 1973, paper TS.1
- [26] A.J.W. Van den BOOM "The determination of the order of process and noise dynamics"
A.W.M. Van den ENDEN
Congrès IFAC La Haye 1973, paper T.S.2
- [27] E. IRVING "Problèmes d'incertitude en identification"
Colloque Franco-Allemand sur les applications industrielles des systèmes adaptatifs. Avril 1973, Fribourg.
- [28] I. GUSTAVSSON "Comparisons of different methods for identification of industrial processes"
Automatica, Vol. 8, pp. 127-142, 1972.
- [29] S. GENTIL "Different methods for dynamic identification of an experimental paper machine"
J.P. SANDRAZ
C. FOULARD
Congrès IFAC La Haye, 1973, paper PV.1

- [30] K.EKLUND
I. GUSTAVSSON "Identification of drum boiler dynamics"
Congrès IFAC, La Haye, 1973, paper C1.
- [31] J.R. WESLAKE "A handbook of numerical matrix inversion and
solution of linear equations"
J. Wiley, 1968.
- [32] C. ECKART
G. YOUNG "The approximation of one matrix by another of
lower rang"
Psychometrika 1, pp. 211-218, 1936.
- [33] P.A. BUSINGER
G.H. GOLUB "Algorithm 358. Singular decomposition of a complex
matrix"
C.A.C.M. Vol. 12, n° 10, october 1969.
- [34] K. FURUTA "An application of realization theory to identifica-
tion of multivariable process"
Congrès IFAC, La Haye, 1973, paper TS-3.
- [35] K.J. ASTROM
T. BOHLIN
S. WENSMARK "Automatic construction of linear stochastic dynamic
models for stationary industrial processes with
random disturbances using operating records"
T.P. 18.150 June 1, 1965, I.B.M. Nordic Laboratory
Sweden.
- [36] D.Q. MAYNE "A canonical model for identification of multivaria-
ble linear system"
I.E.E.E. Trans. Aut. Cont. Vol. AC 17-5 oct. 1972,
pp. 728-729.
- [37] J.D. APLEVICH "Approximation of discrete linear systems"
Int. J. Control, 1973, Vol. 17, n° 3, pp. 565-575.
- [38] J.D. APLEVICH "Gradient methods for optimal linear system reduction"
Int. J. Control, 1973, Vol. 18, n° 4, pp. 767-
- [39] F.D. GALIANA "On the approximation of multiple input multiple
output constant linear systems"
Int. J. Control 1973, Vol. 17, n° 6, pp. 1313-1324.
- [40] R. FLETCHER "A FORTRAN subroutine for minimization by the method
of conjugate gradients"
A.E.R.E. R. 7073, U.K.A.E.A., Harwell, march 1972.
- [41] S. GENTIL
C. FOULARD "Comparaison des résultats de l'identification dynami-
que d'une colonne à distiller pilote, à l'aide de dif-
férentes méthodes statistiques".
RAIRO, série Jaune, Avril 1974, pp. 47-63.
- [42] M.A. MALCOLM "On occurate floating point summation"
C.A.C.M. Vol. 14, n° 11, pp. 731-736 (1971).
- [43] C.L. LAWSON
R.J. HANSON "Solving least squares problems"
Prentice Hall, Inc. 1974.
- [44] Ph. de LARMINAT "L'intérêt et les méthodes de l'identification en vue
de la conduite des systèmes dynamiques"
Journées sur : La simulation en Automatique
8-10 octobre 1974 - E.N.S.M. Nantes

C O N C L U S I O N

Ce mémoire, consacré à la modélisation des systèmes multivariables sous forme d'équations d'état, traite donc de ce qu'on appelle habituellement le problème de *réalisation*, pour des systèmes exactement modélisables non perturbés, et plus généralement d'*approximation* dans le contexte de l'identification de processus physiques réels.

Les travaux effectués depuis une dizaine d'années, essentiellement dans le domaine de la *réalisation minimale*, ont fait apparaître une évolution significative vers l'élaboration de modèles ayant une structure canonique. Cette tendance a défini, en fait, le thème central de notre contribution, développée principalement au Chapitre IV dans le cadre de la réalisation exacte d'un système et au Chapitre V, en ce qui concerne le problème d'approximation.

Avant de parvenir à ces résultats, nous avons rappelé, au cours des deux premiers chapitres quelques points théoriques relatifs aux systèmes linéaires invariants. Il nous a été permis, à cette occasion, d'insister sur certains concepts (accessibilité, commandabilité et leurs duaux observabilité et restructurabilité) dont on parle généralement peu dans le cas discret où se situe notre travail.

Vient ensuite un chapitre de synthèse traitant, d'un type particulier de formes canoniques mono et multivariables, exposé sur la base de la notion d'invariant. Parmi les solutions que l'on pouvait envisager, celle que nous avons retenue (matrice-blocs-compagnons-couplés-dans-un-seul-sens) semble définir une paramétrisation des systèmes, particulièrement bien adaptée à leur identification.

En effet, c'est cette structure même qui a permis de développer des algorithmes effectuant un traitement optimal des données, *sortie par sortie* dans le contexte de la réalisation exacte d'un système (Chapitre IV). Il convient de rappeler ici l'*unicité* du principe suivant lequel ont été déterminées les différentes formes canoniques minimales, du type "accessible" ou "observable", quelle que soit la nature des données à traiter : matrice de transfert en z , séquences de pondération, ou même, et surtout, couples d'entrées/sorties quelconques.

Ce sont ces résultats que nous avons extrapolés, en abordant la modélisation *approchée* d'un système (Chapitre V), domaine qui soulève de nouveaux problèmes.

Une des premières difficultés concernait la définition d'un critère d'approximation. Nous avons retenu deux approches, individuellement bien connues: l'*erreur d'équation* et l'*erreur de sortie*, dont l'utilisation conjointe a caractérisé une procédure hybride exploitant leurs avantages respectifs. Sur le plan de la mise en oeuvre, la minimisation de l'erreur d'équation a été très efficacement obtenue en utilisant les propriétés des transformations de Householder. L'optimisation du second critère, relevant de la programmation non linéaire, a été quant à elle beaucoup plus délicate. Les comparaisons numériques effectuées sur ce point (Chapitre V, § 3) ont clairement mis en évidence la nécessité de faire appel à des algorithmes particulièrement élaborés. Néanmoins là, comme dans beaucoup d'autres applications, il est difficile de définir, a priori, la meilleure politique à suivre. La technique du critère projeté en est un exemple. Appliquée, pensons nous pour la première fois en identification, cette approche ~~permet~~ de réduire la dimension du problème à résoudre à l'ordre du système considéré. On a alors constaté que, bien que cette technique s'avère parfois moins avantageuse qu'un traitement classique du critère, elle offre néanmoins, dans des cas difficiles, un gain de temps calcul qui peut être considérable.

Par ailleurs, ces techniques développées d'abord sur la base d'une formulation monovariable, pour des raisons de simplicité, ont pu être transposées immédiatement aux systèmes multivariables, précisément grâce à la structure particulière des formes canoniques que nous avons retenues. Cependant cette extension qui, d'un point de vue formel, n'offrait aucune difficulté, ne conserve pas, comme on a pu le voir, le caractère optimal que pouvaient présenter, en tant qu'approximation, les résultats obtenus dans le cas monosortie. On a montré, en effet, qu'une approche rigoureuse du problème d'approximation multivariable passait nécessairement par un traitement simultané des sorties.

C'est ce point qui fait ici la différence essentielle entre réalisation et approximation multivariable et qui rend, par voie de conséquence, ce dernier problème beaucoup plus difficile à appréhender. Son ampleur qui nécessiterait des moyens et des méthodes de calculs entrêmement performants explique le peu de tentatives qui ont été faites dans ce domaine.

Une solution effective d'un tel problème se situe, à notre sens, dans la recherche de méthodes sous optimales réalisant le meilleur compromis entre

l'objectif à atteindre et le coût des moyens pour y parvenir. La voie que nous avons retenue consiste à conserver la structure déterminée au cours d'un traitement séquentiel des sorties, puis de compenser le caractère sous optimal de cette première phase par une seconde réalisant une minimisation globale approchée.

L'expérience est là pour confirmer le bien fondé d'une telle démarche qui devrait s'imposer, parmi les techniques d'identification multivariables, pour résoudre des problèmes de dimension réaliste.

Nous voudrions enfin rappeler que le succès même de cette approche a reposé pour une grande part, sur l'importance que l'on a donné à l'aspect numérique du problème. Si celui-ci n'apparaît pas toujours explicitement au cours de ce mémoire, il n'en est pas moins présent dans les résultats numériques que nous avons rapportés et qui témoignent de l'efficacité des algorithmes que nous avons développés.

